

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:35:29 ; Search time 723.713 Seconds  
(without alignments)  
1865.405 Million cell updates/sec

Title: US-09-594-065-4

Perfect score: 33

Sequence: 1 cccgcgaactcgcgtccctccagatgcaggg 33

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues 5777422

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBdb1.\*

1: gb\_da.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pac.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_da.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pac.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rnd.\*

36: em\_htg\_mam.\*

37: em\_htg\_vtc.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.6	71.5	180155	10 AC112273	AC112273 Mus muscu
2	22.6	68.5	2712	6 BD165793	BD165793 Vaccine f
3	22.6	68.5	2715	6 BD145143	BD145143 Herpes si
4	22.6	68.5	2818	14 AF021340	AF021340 Human her
5	22.6	68.5	2846	14 HHU12172	HHU12172 Human herpe
6	22.6	68.5	2846	14 HHU12174	HHU12174 Human herpe
7	22.6	68.5	2855	14 HHU12173	HHU12173 Human herpe
8	22.6	68.5	2855	14 HHU12175	HHU12175 Human herpe
9	22.6	68.5	3324	14 HS2GB	M14923 HSV-2 glycop
10	22.6	68.5	3472	6 BD165792	BD165792 Vaccine f
11	22.6	68.5	3472	6 108685	108685 Sequence 1
12	22.6	68.5	3472	6 HS2GB3	M15118 HSV-2 (333)
13	22.6	68.5	3715	14 HS2GB2A	M24771 Herpes simp
14	22.6	68.5	154746	14 HSV2HG52	286099 Herpes simp
15	22.4	67.9	2281	9 AF181264	AF181264 Homo sapi
16	22.4	67.9	2815	6 AR308082	AR308082 Sequence
17	22.4	67.9	3284	9 BC033100	BC033100 Homo sapi
18	22.4	67.9	3583	9 AF214736	AF214736 Homo sapi
19	22.4	67.9	151949	9 CNS01DS7	AL121657 BAC sequ
20	21.8	66.1	113975	8 OSJN00141	AL662945 Oryza sat
21	21.8	66.1	208920	2 AC140469	AC140469 Mus muscu
22	21.8	66.1	212390	2 AC133576	AC133576 Mus muscu
23	21.4	64.8	798	6 103657	103657 Sequence 5
24	21.4	64.8	1478	14 S74390	S65444 UL27-glycop
25	21.4	64.8	1488	14 S65444	S65444 UL27-glycop
26	21.4	64.8	1857	14 HS1GBM	K02720 HSV1 (mutan
27	21.4	64.8	2088	6 E03093	E03093 DNA encodin
28	21.4	64.8	2712	6 BD145144	BD145144 Herpes si
29	21.4	64.8	2712	6 BD165795	BD165795 Vaccine f
30	21.4	64.8	2715	14 AF097023	AF097023 Human her
31	21.4	64.8	2715	14 AF311740	AF311740 Human her
32	21.4	64.8	2715	14 HHU49121	U49121 Human herpe
33	21.4	64.8	2809	14 AF259899	AF259899 Human her
34	21.4	64.8	2925	6 E03115	E03115 DNA encodin
35	21.4	64.8	3098	6 E00883	E00883 Fragment of
36	21.4	64.8	3461	6 E01195	E01195 DNA sequenc
37	21.4	64.8	3465	6 E03024	E03024 DNA encodin
38	21.4	64.8	3465	6 E03092	E03092 DNA encodin
39	21.4	64.8	3465	6 E03112	E03112 DNA encodin
40	21.4	64.8	3472	6 BD165794	BD165794 Vaccine f
41	21.4	64.8	3472	6 108686	108686 Sequence 2
42	21.4	64.8	3643	6 108361	108361 Sequence 1
43	21.4	64.8	3643	14 HS1GBP	K03541 HSV-1 (Pac
44	21.4	64.8	3755	6 E03025	E03025 DNA encodin
45	21.4	64.8	3755	6 E03113	E03113 DNA encodin

## ALIGNMENTS

RESULT 1

AC112273 180155 bp DNA linear ROD 29-AUG-2002

LOCUS Mus musculus chromosome 3 clone RP24-122H5, complete sequence.

DEFINITION AC112273

ACCESSION AC112273

VERSION AC112273.4 GI:22539282

KEYWORDS HTG.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 180155)

AUTHORS McPherson,J.D. and Waterson,R.H.

TITLE The sequence of Mus musculus clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 180155)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

REFERENCE 3 (bases 1 to 180155)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 180155)  
AUTHORS McPherson,J.D. and Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA

COMMENT On Aug 29, 2002 this sequence version replaced gi:21240758.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@watson.wustl.edu  
----- Project Information -----  
Center project name: W\_B80122405  
-----  
Location/Qualifiers  
1. 180155  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="3"  
/clone="RP24-122H5"

BASE COUNT 51879 a 36032 c 37337 g 54907 t

ORIGIN

Query Match 71.5%; Score 23.6; DB 10; Length 180155;  
Best Local Similarity 81.2%; Pred. No. 29;  
Matches 26; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 CCCTGCAAACTGCTGCTCTCCAGCATGCGAG 32  
|||||  
Db 103608 CCCAGCAACTGCTGCTGCTACACAACTGCGAG 103639

RESULT 2  
BD165793/c  
LOCUS BD165793 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Vaccine for use in the therapeutic treatment of HSV.  
ACCESSION BD165793  
VERSION BD165793.1 GI:27871605  
KEYWORDS JP 2002167398-A/2.  
SOURCE Herpes simplex virus 7  
ORGANISM Herpes simplex virus 7  
Virus;; dDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
1 (bases 1 to 2712)  
Burke,R.L., Pachl,C. and Valenzuela,P.D.T.  
Vaccine for use in the therapeutic treatment of HSV  
Patent: JP 2002167398-A 2 11-JUN-2002;  
CHIRON CORP

COMMENT OS Herpes simplex virus 7  
PN JP 2002167398-A/2  
PD 11-JUN-2002  
PF 12-SEP-2001 JP 2001277340  
PR 20-OCT-1986 US 921.213  
PI RAE LYN BURKE,CAROL PACHL,PABLO D T VALENZUELA PC  
C07K14/035,A61K39/245,A61P31/22,C12N15/09,C12N15/00, PC  
C12N15/00  
CC Vaccine for use in the therapeutic treatment of HSV FH Key  
Location/Qualifiers  
FT CDS  
Location/Qualifiers  
1. 2712

FEATURES  
SOURCE

/organism="Herpes simplex virus 7"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:57277"

BASE COUNT 490 a 938 c 884 g 400 t

ORIGIN

Query Match 68.5%; Score 22.6; DB 6; Length 2712;  
Best Local Similarity 95.7%; Pred. No. 90;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 CAAACTGCTGCTCTCCAGCATG 28  
|||||  
Db 2041 CAAACTGCTGCTCTCCAGCATG 2019

RESULT 3  
BD145143/c  
LOCUS BD145143 2715 bp DNA linear PAT 17-JAN-2003  
DEFINITION Herpes simplex virus VP16 vaccine.  
ACCESSION BD145143  
VERSION BD145143.1 GI:27850901  
KEYWORDS JP 2002136297-A/3.  
SOURCE unidentified  
ORGANISM unidentified  
unclassified.

REFERENCE 1 (bases 1 to 2715)  
AUTHORS Burke,R.L. and Sekulovich,R.E.  
TITLE Herpes simplex virus VP16 vaccine  
JOURNAL Patent: JP 2002136297-A 3 14-MAY-2002;  
CHIRON CORP

COMMENT OS HSV GB2  
PN JP 2002136297-A/3  
PD 14-MAY-2002  
PF 10-SEP-2001 JP 2001274335  
PR 02-AUG-1990 US 561.528  
PI RAE LYN BURKE,ROSE E SEKULOVICH  
PC C12N15/09,A61K39/00,A61K39/245,A61P31/12,C07K14/03,C12N15/00,  
PC A61K37/02  
CC Herpes simplex virus VP16 vaccine  
FH Key  
Location/Qualifiers  
FT CDS  
Location/Qualifiers  
1. 2715  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 491 a 938 c 886 g 400 t

ORIGIN

Query Match 68.5%; Score 22.6; DB 6; Length 2715;  
Best Local Similarity 95.7%; Pred. No. 90;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 CAAACTGCTGCTCTCCAGCATG 28  
|||||  
Db 2041 CAAACTGCTGCTCTCCAGCATG 2019

RESULT 4  
AF021340/c  
LOCUS AF021340 2818 bp DNA linear VRL 24-AUG-2000  
DEFINITION Human herpesvirus 2 glycoprotein B precursor (UL27) gene, complete  
cde.  
ACCESSION AF021340  
VERSION AF021340.1 GI:2465440  
KEYWORDS Human herpesvirus 2  
SOURCE Human herpesvirus 2  
ORGANISM Human herpesvirus 2  
Virus;; dDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
1 (bases 1 to 2818)  
Tehrane,S.S., Coleman,K.T., Sekulovich,R., Burke,R.L. and  
Spear,P.G.

REFERENCE 1 (bases 1 to 2818)  
AUTHORS

**TITLE** Limited variability of glycoprotein gene sequences and neutralizing passages in herpes simplex virus type 2 isolates and stability on passage in cell culture

**JOURNAL** J. Infect. Dis. 178 (1), 8-15 (1998)

**MEDLINE** 98314525

**PUBMED** 9652417

**REFERENCE** 2 (bases 1 to 2818)

**AUTHORS** Terhune, S.S., Coleman, K.T., Sekulovich, R., Burke, R.L. and Spear, P.G.

**TITLE** Direct Submission

**JOURNAL** Submitted (29-AUG-1997) Microbiology-Immunology, Northwestern University, 303 East Chicago Ave, Chicago, IL 60611, USA

**FEATURES** Location/Qualifiers

**source** 1..2818

**organism**="Human herpesvirus 2"

**/mol\_type**="genomic DNA"

**/strain**="BHKC"

**/db\_xref**="taxon:10310"

**gene** 1..2715

**/gene**="UL27"

**/note**="UL27"

**/codon\_start**=1

**/product**="glycoprotein B precursor"

**/protein\_id**="AAB72100.1"

**/db\_xref**="GI:2465441"

**/translation**="MRGGGLICALVGVALVAASAPAPAPRAGGVATVAANG  
GPARPPVPSPTATRAKRTKPKPRPEATPPDANATVAAGHATLRHLEIKYE  
MAAOFPVCPPTGATVVOPEORPCPTREGNTGIAVVKENIAPKPKATRYKY  
KDTVSQVWGHRYSQPMGI FEDAPVPFEVIDIKINAKGVCS TAKYVNNMETTA  
HRDHEITDMLKPAKVAFTSRGSHHTTDLKNSRVAEPHRYGTTVNCIVEEDARS  
YPDEFLATGDPVYMSFPYGRSGHTEHTSYAARFKVDGFYADLTITKAQATSP  
TTRILTLTPKFTVAMWVPKRPACVTMTKQEVDEMILBAEYGSFRSSDAISTFTT  
NLQYLSRVLDGDCIGDRAEALIDMPAKVNAITHIKVGQPOVYLA TGGLIAYOPL  
LSHTALVYREYREBDKRPNAITPAPLRASANSVERIKTSSIEPARLOFTYN  
HIOHVDMLGRILAVANCELOHNELTLMNEARKLPALASATVGRVSRMLGDAVST  
VSTCVPAEDNVIVONSMRVSSRPCTYSRPLVSFRYEDGPILEGQGENNELRLR  
DALEPCTVGHRRYFPIGGGVYFEERYASHQLSRADVTYVSTFDLNIITMLEDEFPV  
LEVYTRHEIDSGILDTYEVORNOHLDRFADIDTVIRADANAMFAGCAFEFGG  
DLGAVGVGVGVAVSAGVSSFMSPFGALVAGLVLAGLVAAFPAPFRVYLOLN  
QRNPKALVPLTTELTSDPGYGVGBEGGAGCGGDEKALBAREMITMALVSM  
ERHEKARKGTGTSALLSKVTNVLKRNKARYSPHNEDAGDEDEL"

**BASE COUNT** 515 a 952 c 920 g 431 t

**ORIGIN**

**Query Match** 68.5%; Score 22.6; DB 14; Length 2818;

**Best Local Similarity** 95.7%; Pred. No. 90;

**Matches** 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

**Cy** 6 CAAACTGCTGCTCTCCAGCATG 28

**Db** 2041 CAAACTGCTGCTCTCCAGCATG 2019

**RESULT 5** HHU12172 2846 bp DNA linear VRL 05-AUG-1994

**LOCUS** Human herpesvirus 2 CAM4B glycoprotein B (UL27) gene, complete cds.

**DEFINITION** U12172

**ACCESSION** U12172

**VERSION** U12172.1 GI:517447

**KEYWORDS** Human herpesvirus 2

**SOURCE** Human herpesvirus 2

**ORGANISM** Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Simplexvirus.

**REFERENCE** 1 (bases 1 to 2846)

**AUTHORS** Terhune, S.S. and Spear, P.G.

**TITLE** Variability of herpes simplex virus glycoproteins in clinical specimens

**JOURNAL** Unpublished

**REFERENCE** 2 (bases 274 to 2846)

**AUTHORS** Stuve, L.L., Brown-Shimer, S., Pachl, C., Najarian, R., Dina, D. and Burke, R.L.

**TITLE** Structure and expression of the herpes simplex virus type 2 glycoprotein gB gene

**JOURNAL** J. Virol. 61 (2), 326-335 (1987)

**MEDLINE** 87112925

**PUBMED** 3027364

**REFERENCE** 3 (bases 1 to 2846)

**AUTHORS** Spear, P.G.

**TITLE** Direct Submission

**JOURNAL** Submitted (11-JUL-1994) Patricia G. Spear, Microbiology-Immunology, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, IL 60611, USA

**FEATURES** Location/Qualifiers

**source** 1..2846

**organism**="Human herpesvirus 2"

**/mol\_type**="genomic DNA"

**/strain**="CAM4B"

**/db\_xref**="taxon:10310"

**gene** 36..2741

**/gene**="UL27"

**/note**="gB; precursor peptide"

**/codon\_start**=1

**/product**="glycoprotein B"

**/protein\_id**="AAB60545.1"

**/db\_xref**="GI:522168"

**/translation**="MRGGGLICALVGVALVAASAPAPRAGGVATVAANGPA  
SOPRPVPSPTATRAKRTKPKPRPEATPPDANATVAAGHATLRHLEIKYEND  
AOYVCPPTGATVVOPEORPCPTREGNTGIAVVKENIAPKPKATRYKY  
TVSQVWGHRYSQPMGI FEDAPVPFEVIDIKINAKGVCS TAKYVNNMETTA  
HRDHEITDMLKPAKVAFTSRGSHHTTDLKNSRVAEPHRYGTTVNCIVEEDARS  
YPDEFLATGDPVYMSFPYGRSGHTEHTSYAARFKVDGFYADLTITKAQATSP  
TTRILTLTPKFTVAMWVPKRPACVTMTKQEVDEMILBAEYGSFRSSDAISTFTT  
NLQYLSRVLDGDCIGDRAEALIDMPAKVNAITHIKVGQPOVYLA TGGLIAYOPL  
LSHTALVYREYREBDKRPNAITPAPLRASANSVERIKTSSIEPARLOFTYN  
HIOHVDMLGRILAVANCELOHNELTLMNEARKLPALASATVGRVSRMLGDAVST  
VSTCVPAEDNVIVONSMRVSSRPCTYSRPLVSFRYEDGPILEGQGENNELRLR  
DALEPCTVGHRRYFPIGGGVYFEERYASHQLSRADVTYVSTFDLNIITMLEDEFPV  
LEVYTRHEIDSGILDTYEVORNOHLDRFADIDTVIRADANAMFAGCAFEFGGDLG  
AVGVGVGVAVSAGVSSFMSPFGALVAGLVLAGLVAAFPAPFRVYLOLN  
QRNPKALVPLTTELTSDPGYGVGBEGGAGCGGDEKALBAREMITMALVSMERT  
ERHEKARKGTGTSALLSKVTNVLKRNKARYSPHNEDAGDEDEL"

**BASE COUNT** 522 a 962 c 930 g 432 t

**ORIGIN**

**Query Match** 68.5%; Score 22.6; DB 14; Length 2846;

**Best Local Similarity** 95.7%; Pred. No. 90;

**variation** 172

**/gene**="UL27"

**/citation**=(2)

**/replace**="cgsgcgcccc"

**variation** 205

**/gene**="UL27"

**/citation**=(2)

**/replace**="a"

**variation** 876

**/gene**="UL27"

**/citation**=(2)

**/replace**="t"

**variation** 1015

**/gene**="UL27"

**/citation**=(2)

**/replace**="g"

**variation** 1250

**/gene**="UL27"

**/citation**=(2)

**/replace**="t"

Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 6 CAAACTCGTGTCTCCAGCATG 28  
 |||||  
 Db 2067 CAAACTCGTGTCTCCAGCATG 2045

RESULT 6  
 HHU12174/c 2846 bp DNA linear VRL 05-AUG-1994

LOCUS Human herpesvirus 2 MMA glycoprotein B (UL27) gene, complete cds.

DEFINITION U12174

ACCESSION U12174.1 GI:517451

VERSION U12174.1

KEYWORDS Human herpesvirus 2

SOURCE Human herpesvirus 2

ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

REFERENCE 1 (bases 1 to 2846)

AUTHORS Terhune, S.S. and Spear, P.G.

TITLE Variability of herpes simplex virus glycoproteins in clinical specimens

JOURNAL Unpublished

REFERENCE 2 (bases 274 to 2846)

AUTHORS Stuve, L.L., Brown-Shimer, S., Pachl, C., Najarian, R., Dina, D. and Burke, R.L.

TITLE Structure and expression of the herpes simplex virus type 2 glycoprotein gB gene

JOURNAL J. Virol. 61 (2), 326-335 (1987)

REFERENCE 3 (bases 1 to 2846)

AUTHORS Spear, P.G.

TITLE Direct Submision

REFERENCE Submitted (11-JUL-1994) Patricia G. Spear, Microbiology-Immunology,

Northwestern University Medical School, 303 East Chicago Avenue,

Chicago, IL 60611, USA

LOCATION/Qualifiers

1. 2846

/organism="Human herpesvirus 2"

/mol\_type="genomic DNA"

/strain="MMA"

/db\_xref="taxon:10310"

36..2741

/gene="UL27"

36..2741

/gene="UL27"

/note="gB; precursor peptide"

/codon\_start=1

/product="glycoprotein B"

/protein\_id="AA060547.1"

/db\_xref="GI:522170"

/translation="MRGGGLCALVYGVLAANVAASAPAPASGAGVATYAAAGPRA

SRPPVSPATIRAKRKTKKPEKPEKATPPDAATVAAGATATRAHLEIKENAD

AOFYVCPPTGATVQFEQPRCPTRPGQVYTBELIAVFENAIAPYFKATMYKDV

TVSGVQFGRYSQFWGIFEDRPAVFEVEIKINAKGCRSTAKYRNMMETTFHRD

DHEITMELKPAKVAATRTSGMTTDLKYNPGRVAFHNGTTCVIEVDARSVPY

DEPLVATGDFVYMSPPYGRSGSHTEHSVAADRFQVGFARLITKAAATSPTR

NLTTPKRTVAMDWPKRPAYCTMKQENDEM.PAYGSGSPRSSDASTPTTNLT

EYSLSRDLGCTIGDAREATDRMPARKYMTHTKVGOPYLATGFLIYOPLLSN

TLAELVREYREEDRKRNATIPALREAPASMAVERIKTTSIEFARLQFTNNHQ

RHVNDMLGRIVAMCELONHETLMMENARKLPNAIASATVGRVSAMLQDVAVST

CVVAPDNVIVONSMRVSSRPGTCYSRLPLVFRYDQPLIEGQGENNELRLTRDAL

EPCTVGHRYPIFGGYYFBEYVAASHOLSHADVTVTFTDINLTMLDEHPPLVY

YTRHETKSGSLDTEVORRNQHLRADIITYIRANAMAPGLCAFPFGQCDLG

RAVGKRVGGVGVSAVSGVSSFNPFGLAVGLVLAGLVAAFFAFRYVLQDRN

EMKALYPLTKELKTSIDPGVGVGEGEBAEGGPEALAEAREMIRYMALVAMERT

EHKARKGTSALSSKVTNMYLARKNKARYSPLFHNEDAGDEDEL"

36..101

/gene="UL27"

97

/gene="UL27"

/citation=[2]

variation /replace="cgsgsgcccc"

143

/gene="UL27"

/citation=[2]

/replace="c"

205

/gene="UL27"

/citation=[2]

/replace="a"

237

/gene="UL27"

/citation=[2]

/replace="a"

876

/gene="UL27"

/citation=[2]

/replace="c"

1015

/gene="UL27"

/citation=[2]

/replace="g"

1212

/gene="UL27"

/citation=[2]

/replace="c"

1250

/gene="UL27"

/citation=[2]

/replace="c"

variation

variation

variation

variation

variation

variation

variation

BASE COUNT 520 a 959 c 934 g 433 t

ORIGIN

Query Match 68.5%; Score 22.6; DB 14; Length 2846;

Best Local Similarity 95.7%; Pred. No. 90;

Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 6 CAAACTCGTGTCTCCAGCATG 28

|||||

Db 2067 CAAACTCGTGTCTCCAGCATG 2045

|||||

RESULT 7

HHU12173/c 2855 bp DNA linear VRL 05-AUG-1994

LOCUS Human herpesvirus 2 JD23 glycoprotein B (UL27) gene, complete cds.

DEFINITION U12173

ACCESSION U12173.1 GI:517449

VERSION U12173.1

KEYWORDS Human herpesvirus 2

SOURCE Human herpesvirus 2

ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

REFERENCE 1 (bases 1 to 2855)

AUTHORS Terhune, S.S. and Spear, P.G.

TITLE Variability of herpes simplex virus glycoproteins in clinical specimens

JOURNAL Unpublished

REFERENCE 2 (bases 274 to 2855)

AUTHORS Stuve, L.L., Brown-Shimer, S., Pachl, C., Najarian, R., Dina, D. and Burke, R.L.

TITLE Structure and expression of the herpes simplex virus type 2 glycoprotein gB gene

JOURNAL J. Virol. 61 (2), 326-335 (1987)

REFERENCE 3 (bases 1 to 2855)

AUTHORS Spear, P.G.

TITLE Direct Submision

REFERENCE Submitted (11-JUL-1994) Patricia G. Spear, Microbiology-Immunology,

Northwestern University Medical School, 303 East Chicago Avenue,

Chicago, IL 60611, USA

LOCATION/Qualifiers

1. 2855

/organism="Human herpesvirus 2"

/mol\_type="genomic DNA"

/strain="MMA"

/db\_xref="taxon:10310"

36..2741

/gene="UL27"

36..2741

/gene="UL27"

/note="gB; precursor peptide"

/codon\_start=1

/product="glycoprotein B"

/protein\_id="AA060547.1"

/db\_xref="GI:522170"

/translation="MRGGGLCALVYGVLAANVAASAPAPASGAGVATYAAAGPRA

SRPPVSPATIRAKRKTKKPEKPEKATPPDAATVAAGATATRAHLEIKENAD

AOFYVCPPTGATVQFEQPRCPTRPGQVYTBELIAVFENAIAPYFKATMYKDV

TVSGVQFGRYSQFWGIFEDRPAVFEVEIKINAKGCRSTAKYRNMMETTFHRD

DHEITMELKPAKVAATRTSGMTTDLKYNPGRVAFHNGTTCVIEVDARSVPY

DEPLVATGDFVYMSPPYGRSGSHTEHSVAADRFQVGFARLITKAAATSPTR

NLTTPKRTVAMDWPKRPAYCTMKQENDEM.PAYGSGSPRSSDASTPTTNLT

EYSLSRDLGCTIGDAREATDRMPARKYMTHTKVGOPYLATGFLIYOPLLSN

TLAELVREYREEDRKRNATIPALREAPASMAVERIKTTSIEFARLQFTNNHQ

RHVNDMLGRIVAMCELONHETLMMENARKLPNAIASATVGRVSAMLQDVAVST

CVVAPDNVIVONSMRVSSRPGTCYSRLPLVFRYDQPLIEGQGENNELRLTRDAL

EPCTVGHRYPIFGGYYFBEYVAASHOLSHADVTVTFTDINLTMLDEHPPLVY

YTRHETKSGSLDTEVORRNQHLRADIITYIRANAMAPGLCAFPFGQCDLG

RAVGKRVGGVGVSAVSGVSSFNPFGLAVGLVLAGLVAAFFAFRYVLQDRN

EMKALYPLTKELKTSIDPGVGVGEGEBAEGGPEALAEAREMIRYMALVAMERT

EHKARKGTSALSSKVTNMYLARKNKARYSPLFHNEDAGDEDEL"

36..101

/gene="UL27"

97

/gene="UL27"

/citation=[2]



Northwestern University Medical School, 303 East Chicago Avenue,  
Chicago, IL 60611, USA

## FEATURES

## SOURCE

1. .2855

/organism="Human herpesvirus 2"

/mol\_type="genomic DNA"

/strain="JD23"

/db\_xref="taxon:10310"

36. .2750

/gene="UL27"

36. .2750

/note="gb: precursor peptide"

/codon\_start=1

/product="glycoprotein B"

/protein\_id="AA60546.1"

/db\_xref="GI:522169"

/translation="MRGGSLICLVGALVAASAPAAPAPRASGVAAATYANG  
GPASRPPVSPATTKARKKTKKPPRPPATPPDANATVAAGHATLRALREIKYE  
NADAQFYVCPPTGATVVOEOPRCPTREGQNYTEGIAVFEKNTAPYFKATMY  
KDVTSQVMFGHRSQFMGI FEDRAPVPFEVIDIKNAKGVCRSTACYVRNNMTAF  
HRDDETMDLKPAPKAVATRTSRGHTTDLKNSRVPAPFRYGTTCVIEVDARSV  
YPRDEFLATGDPVYNSPFYGRGSGHTHTSYAADFKQVDFYARDLTTKARISB  
TTNNLTTPKFTVAMDVPPRPVACTMTKQVDEMTAKETGSGFRSSDAISTFTT  
NLTEYLSRVLDGICIGDARBAIDRMFAKRYNATHIKVQPOYTLATGGFLAYOPL  
LSNTLAEIYVREYREBQDKRPKNATPPALREAPSANASVERIKTSSIEFARLOFTYN  
HIOHVNMDLGRILAVANCELOHNEHLTLMNEARKNPALIASATGRRVSARMGDVWA  
VSTCVPAEDNVIVQNSMRVSRGTCSPVLSFRYEDGSPILIEGOLGNNEELRLR  
DALEPCTVGHRRTYIFGGGYVPEBYASHQLSRADVTSTFTIDLNTLMEDHEFPV  
LEVYTRHEIKDSGLDYTEVQRNQLHDLRFADIDIVIRADANAMAPGACAFPEGNG  
DLGRAVGVVGVGVSASVSSVSSFMSPFGALAVGLVLAGLVAFAPFRVYLQ  
ORNPMLALYPLTTKEIKTSDPGGVGGEAGCGDEAKLAEREMIRMALVSAM  
ERTEHKARKKGTALSSKVTNNVLRKRNKARYSPLEHDEAGDEDEL"

36. .101

/gene="UL27"

885

/gene="UL27"

/citation=[2]

/replace="t"

1221

/gene="UL27"

/citation=[2]

/replace="c"

1259

/gene="UL27"

/citation=[2]

/replace="t"

2282

/gene="UL27"

/citation=[2]

/replace="a"

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Alpha herpesvirinae; Simplexvirus.  
1 (bases 1 to 2855)  
Terhune, S.S. and Spear, P.G.  
Variability of herpes simplex virus glycoproteins in clinical  
specimens  
Unpublished  
2 (bases 274 to 2855)  
Stuve, L.L., Brown-Shimer, S., Pachl, C., Najarian, R., Dina, D. and  
Burke, R.L.  
Structure and expression of the herpes simplex virus type 2  
glycoprotein gB gene  
J. Virol. 61 (2), 326-335 (1987)  
87112925  
3027364  
3 (bases 1 to 2855)  
Spear, P.G.  
Direct Submission  
Submitted (11-JUL-1994) Patricia G. Spear, Microbiology-Immunology,  
Northwestern University Medical School, 303 East Chicago Avenue,  
Chicago, IL 60611, USA

Location/Qualifiers

1. .2855

/organism="Human herpesvirus 2"

/mol\_type="genomic DNA"

/strain="WTMIA"

/db\_xref="taxon:10310"

36. .2750

/gene="UL27"

36. .2750

/note="gb: precursor peptide"

/codon\_start=1

/product="glycoprotein B"

/protein\_id="AA60546.1"

/db\_xref="GI:522171"

/translation="MRGGSLICLVGALVAASAPAAPAPRASGVAAATYANG  
GPASRPPVSPATTKARKKTKKPPRPPATPPDANATVAAGHATLRALREIKYE  
NADAQFYVCPPTGATVVOEOPRCPTREGQNYTEGIAVFEKNTAPYFKATMY  
KDVTSQVMFGHRSQFMGI FEDRAPVPFEVIDIKNAKGVCRSTACYVRNNMTAF  
HRDDETMDLKPAPKAVATRTSRGHTTDLKNSRVPAPFRYGTTCVIEVDARSV  
YPRDEFLATGDPVYNSPFYGRGSGHTHTSYAADFKQVDFYARDLTTKARISB  
TTNNLTTPKFTVAMDVPPRPVACTMTKQVDEMTAKETGSGFRSSDAISTFTT  
NLTEYLSRVLDGICIGDARBAIDRMFAKRYNATHIKVQPOYTLATGGFLAYOPL  
LSNTLAEIYVREYREBQDKRPKNATPPALREAPSANASVERIKTSSIEFARLOFTYN  
HIOHVNMDLGRILAVANCELOHNEHLTLMNEARKNPALIASATGRRVSARMGDVWA  
VSTCVPAEDNVIVQNSMRVSRGTCSPVLSFRYEDGSPILIEGOLGNNEELRLR  
DALEPCTVGHRRTYIFGGGYVPEBYASHQLSRADVTSTFTIDLNTLMEDHEFPV  
LEVYTRHEIKDSGLDYTEVQRNQLHDLRFADIDIVIRADANAMAPGACAFPEGNG  
DLGRAVGVVGVGVSASVSSVSSFMSPFGALAVGLVLAGLVAFAPFRVYLQ  
ORNPMLALYPLTTKEIKTSDPGGVGGEAGCGDEAKLAEREMIRMALVSAM  
ERTEHKARKKGTALSSKVTNNVLRKRNKARYSPLEHDEAGDEDEL"

36. .101

/gene="UL27"

203

/gene="UL27"

/citation=[2]

/replace="c"

885

/gene="UL27"

/citation=[2]

/replace="t"

1221

/gene="UL27"

/citation=[2]

/replace="c"

1259

/gene="UL27"

/citation=[2]

/replace="t"

936 g

433 t

521 a

965 c

936 g

433 t

521 a

965 c

936 g

433 t

521 a

965 c

936 g

433 t

521 a

965 c

936 g

433 t

521 a

965 c

936 g

433 t

521 a

965 c

936 g

433 t

521 a

965 c

## FEATURES

## SOURCE

1. .2855

/organism="Human herpesvirus 2"

/mol\_type="genomic DNA"

/strain="JD23"

/db\_xref="taxon:10310"

36. .2750

/gene="UL27"

36. .2750

/note="gb: precursor peptide"

/codon\_start=1

/product="glycoprotein B"

/protein\_id="AA60546.1"

/db\_xref="GI:522169"

/translation="MRGGSLICLVGALVAASAPAAPAPRASGVAAATYANG  
GPASRPPVSPATTKARKKTKKPPRPPATPPDANATVAAGHATLRALREIKYE  
NADAQFYVCPPTGATVVOEOPRCPTREGQNYTEGIAVFEKNTAPYFKATMY  
KDVTSQVMFGHRSQFMGI FEDRAPVPFEVIDIKNAKGVCRSTACYVRNNMTAF  
HRDDETMDLKPAPKAVATRTSRGHTTDLKNSRVPAPFRYGTTCVIEVDARSV  
YPRDEFLATGDPVYNSPFYGRGSGHTHTSYAADFKQVDFYARDLTTKARISB  
TTNNLTTPKFTVAMDVPPRPVACTMTKQVDEMTAKETGSGFRSSDAISTFTT  
NLTEYLSRVLDGICIGDARBAIDRMFAKRYNATHIKVQPOYTLATGGFLAYOPL  
LSNTLAEIYVREYREBQDKRPKNATPPALREAPSANASVERIKTSSIEFARLOFTYN  
HIOHVNMDLGRILAVANCELOHNEHLTLMNEARKNPALIASATGRRVSARMGDVWA  
VSTCVPAEDNVIVQNSMRVSRGTCSPVLSFRYEDGSPILIEGOLGNNEELRLR  
DALEPCTVGHRRTYIFGGGYVPEBYASHQLSRADVTSTFTIDLNTLMEDHEFPV  
LEVYTRHEIKDSGLDYTEVQRNQLHDLRFADIDIVIRADANAMAPGACAFPEGNG  
DLGRAVGVVGVGVSASVSSVSSFMSPFGALAVGLVLAGLVAFAPFRVYLQ  
ORNPMLALYPLTTKEIKTSDPGGVGGEAGCGDEAKLAEREMIRMALVSAM  
ERTEHKARKKGTALSSKVTNNVLRKRNKARYSPLEHDEAGDEDEL"

36. .101

/gene="UL27"

885

/gene="UL27"

/citation=[2]

/replace="t"

1221

/gene="UL27"

/citation=[2]

/replace="c"

1259

/gene="UL27"

/citation=[2]

/replace="a"

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

520 a

967 c

936 g

432 t

Best Local Similarity 95.7%; Pred. No. 90;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAACTCGTGCTCCAGCATG 28  
|||||  
Db 2076 CAAACTCGTGCTCCAGCATG 2054

RESULT 9  
HS2GB/c 3324 bp DNA linear VRL 15-MAY-1995

LOCUS HS2GB 3324 bp DNA linear VRL 15-MAY-1995  
DEFINITION HS2GB glycoprotein B gene, complete cds.  
ACCESSION M14923  
VERSION M14923.1 GI:330254  
KEYWORDS glycoprotein; glycoprotein gb.  
SOURCE Human herpesvirus 2  
ORGANISM Human herpesvirus 2

REFERENCE 1  
AUTHORS Bzik,D.J., Debroy,C., Fox,B.A., Pederson,N.E. and Person,S.  
TITLE The nucleotide sequence of the gb glycoprotein gene of HSV-2 and comparison with the corresponding gene of HSV-1  
JOURNAL Virology 155 (2), 322-333 (1986)  
MEDLINE 87071654  
PUBMED 3024391

COMMENT Original source text: Herpes simplex virus type 2 (strain HG52)

DNA  
Draft entry and computer-readable sequence of [1] kindly provided by N.E. Pederson, 10-APR-1987.  
The sequences of gb glycoprotein genes of HSV-1 and HSV-2 were compared and there was an overall amino acid homology of 85%. There were three major regions of homology at nucleotides 108-197, 484-785, and 835-873.  
A membrane spanning region was found at position 2717-2923. The authors proposed that a signal cleavage-recognition site exists after nucleotide 613. This would give a mature peptide of 882 residues, from 614-1262.  
Location/Qualifiers

FEATURES  
source 1..3324

CDS 1..3324  
/organism="Human herpesvirus 2"  
/mol\_type="genomic DNA"  
/strain="HG52"  
/db\_xref="taxon:10310"  
/lab\_host="HEL cell"  
/note="ORF, putative"  
/codon\_start=1  
/product="unknown protein"  
/protein\_id="AAA66439.1"  
/db\_xref="GI:807643"  
/translation="VNGPLFDHSTHSPAPPTALYSEVENVGLPHLKEIARPIAGSGADMANVSEFQKFCFDVSGITFQKAPAKRITRLIATITLPAVYGCGELELRKPDSCRPTSEGLRYPGCVLYTNSDCLPVALIVESGPDGIGPRVVVYDVSILYSVLQHLAPRLAGSGDADP"

mRNA 261..3326  
/note="gb mRNA (5' end +/- 4 bp)"  
548..3362  
/codon\_start=1  
/product="glycoprotein gb"  
/protein\_id="AAA66440.1"  
/db\_xref="GI:330255"  
/translation="MRGGGLICALVVALVAASAPAAPAPRAGSGVAVTAANGGASRPVPPPTATTKARKRKTKPKPKPEATPPDANAATVAAAGATVRAIRIKIYENADQFYCCPPPTGATVVOFEQPRACPRPREGQNTTEIAYVRENIAPIYFKATMYKDVTSQVAFGRISQFMGIFEDRAPVPEVEVIDINKAGCRSTAKTVRNMMETIARHDDHETMELMELPAVARTSGMHTTDLKNPSVKAFAHGGITVNCIVEVARSVPYDEFLATGDFVMSPPYGRGESHETTYADREKQYDGFYARDLTAKATSPTRNLTLTPKFTVANDVVKRPVACTMTKQVDEMLBAEYGSFRFSDDAISTFTNUTVSLSRVLDGICIGRDAREADIDRMFARKNAHTIKVGGPOVYQATGGLIAYOPLLSNTLAELIYREYAEQPRKRNATPALREKPSANASVERIKTTSSIEPRALQPTVNLQIRVNMGLGRIAVAWCELOHETLTNNKAKLPNPIASATVGRRVASAMLDGVMAVATCVPAVDNVIYONSMKRVSSRPGTCYSRPLVSFRYEDGPLLEGQLENNVRLTR

DALEPCTVGRGFIIFGGYVFEVAVSHQLSRADVTVTSTPIDLITMLDEHEVPL  
LEVTYRHEIDSGILDYTVORRNQDLHPADIDTVIRADANAMAPGACAPGEMG  
DLGRAVGVGVAVGAVSAVGSFSPNSPFGALVGLVLAAGLVAFAPFRVLOL  
ORNMKALYPLTTKEIKLSDPBGVGEAGEGGEDEAKLAEAREMIRYMALVSAM  
ERTHEKARKGTGALLSKYTNMVLRRNARYSPLINEBAGEDEDEL"

BASE COUNT 600 a 1151 c 1063 g 510 t  
ORIGIN 242 bp upstream of NruI site; between 0.370 and 0.348 map units.

Query Match 68.5%; Score 22.6; DB 14; Length 3324;  
Best Local Similarity 95.7%; Pred. No. 89;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAACTCGTGCTCCAGCATG 28  
|||||  
Db 2588 CAAACTCGTGCTCCAGCATG 2566

RESULT 10  
BD165792/c 3472 bp DNA linear PAT 17-JAN-2003  
LOCUS BD165792  
DEFINITION Vaccine for use in the therapeutic treatment of HSV.  
ACCESSION BD165792  
VERSION BD165792.1 GI:27871604  
KEYWORDS JP 2002167398-A/1.  
SOURCE Herpes simplex virus 7  
ORGANISM Herpes simplex virus 7

REFERENCE 1  
AUTHORS Burke,R.L., Pachl,C. and Valenzuela,P.D.T.  
TITLE Vaccine for use in the therapeutic treatment of HSV  
JOURNAL Patent: JP 2002167398-A 1 11-JUN-2002;  
CHIRON CORP  
COMMENT OS Herpes simplex virus 7  
PN JP 2002167398-A/1  
PD 11-JUN-2002  
PF 12-SEP-2001 JP 2001277340  
PR 20-OCT-1986 US 921,213  
PI RAB LYN BURKE, CAROL PACHL, PABLO D T VALENZUELA PC  
COTK14/035, A61K39/245, A61P31/22, C12N15/09, C12N15/00, C12N15/00  
CC Vaccine for use in the therapeutic treatment of HSV FH Key

FEATURES  
source 1..3472  
/organism="Herpes simplex virus 7"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:57277"

BASE COUNT 639 a 1202 c 1069 g 562 t  
ORIGIN

Query Match 68.5%; Score 22.6; DB 6; Length 3472;  
Best Local Similarity 95.7%; Pred. No. 89;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAACTCGTGCTCCAGCATG 28  
|||||  
Db 2349 CAAACTCGTGCTCCAGCATG 2327

RESULT 11  
108685/c 3472 bp DNA linear PAT 02-DEC-1994  
LOCUS 108685  
DEFINITION Sequence 1 from Patent WO 8802634.  
ACCESSION 108685  
VERSION 108685.1 GI:588605  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 3472)

AUTHORS Burke,R.L., Pachl,C. and Valenzuela,P.D.T.

BASE COUNT  
639 a 1202 c 1069 g 562 t  
3 bp upstream of NruI site; 0.344 to 0.367 mu on HSV-2 map.

YDEFLVATGDFVYMSPPYGRGESHTEHTTYAARFKQVGDGYARDLTTKARATSPPT  
RNLITTPKFTVANDWVPKRPAVCTMTKQEVDEMLRAEYGGSGFRPSSDAISTFTTNL

Query Match	68.5%	Score 22.6	DB 14	Length 3715
Beat local similarity	95.7%	Pred. No. 89		
Matches 22	Conservative 1	Mismatches 0	Indels 0	Gaps 0
Db	2590	CAAACTCGTGGCTCCGACGATG	2568	
RESULT 14				
LOCUS	HSV2HG52	154746 bp	DNA	linear
DEFINITION	Herpes simplex virus type 2 (strain HG52), complete genome.			
ACCESSION	286099			
VERSION	286099.2			
KEYWORDS	capcid associated tegument protein; capcid protein; complete genome; deoxyribonuclease; deoxyridine triphosphatase; DNA polymerase; envelope protein; glycoprotein B; glycoprotein C; glycoprotein D; glycoprotein E; glycoprotein G; glycoprotein H; glycoprotein I; glycoprotein J; glycoprotein K; glycoprotein L; glycoprotein M; host shut-off factor; integral membrane protein; major capid protein; membrane-associated phosphoprotein; minor capid protein; neurovirulence factor; protease; protein kinase; ribonucleotide reductase large subunit; ribonucleotide reductase small subunit; R1 gene; R2 gene; R3 gene; tegument protein; thymidine kinase; UL1 gene; UL10 gene; UL11 gene; UL12 gene; UL13 gene; UL14 gene; UL15 gene; UL16 gene; UL17 gene; UL18 gene; UL19 gene; UL20 gene; UL21 gene; UL22 gene; UL23 gene; UL24 gene; UL25 gene; UL26 gene; UL26.5 gene; UL27 gene; UL28 gene; UL29 gene; UL30 gene; UL31 gene; UL32 gene; UL33 gene; UL34 gene; UL35 gene; UL36 gene; UL37 gene; UL38 gene; UL39 gene; UL40 gene; UL41 gene; UL42 gene; UL43 gene; UL44 gene; UL45 gene; UL46 gene; UL47 gene; UL48 gene; UL49 gene; UL49A gene; UL50 gene; UL51 gene; UL52 gene; UL53 gene; UL54 gene; UL55 gene; UL56 gene; UL6 gene; UL7 gene; UL8 gene; UL9 gene; uracil-DNA glycosylase; US1 gene; US10 gene; US11 gene; US12 gene; US2 gene; US3 gene; US4 gene; US5 gene; US6 gene; US7 gene; US8 gene; US9A gene; US9 gene.			
SOURCE	Human herpesvirus 2			
ORGANISM	Human herpesvirus 2			
REFERENCE	Alphaherpesvirinae, no RNA stage; Herpesviridae; Alphaherpesvirinae; Simplexvirus.			
AUTHORS	McGeoch, D. J., Moss, H. M., McNab, D. and Pringle, M. C.			
TITLE	DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons			
JOURNAL	J. Gen. Virol.	68 (Pt 1),	19-38	(1987)
MEDLINE	87111457			
PubMed	3027242			
REFERENCES	2 (bases 1 to 154746)			
AUTHORS	Everett, R. D. and Fenwick, M. L.			

REFERENCE	AUTHORS	TITLE
JOURNAL	McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.	Comparative DNA sequence analysis of the host shutoff genes of different strains of herpes simplex virus: type 2 strain H052
MEDLINE	J. Gen. Virol. 71 (Pt 6), 1387-1390 (1990)	encodes a truncated UL41 product
PUBMED	90278430	
REFERENCE	3 (bases 1 to 154746)	
AUTHORS	McGeoch,D.J., Cunningham,C., McIntyre,G. and Dolan,A.	Comparative sequence analysis of the long repeat regions and adjoining parts of the long unique regions in the genomes of herpes simplex viruses types 1 and 2
TITLE	J. Gen. Virol. 72 (Pt 12), 3057-3075 (1991)	
JOURNAL	92113549	
MEDLINE	1662697	
PUBMED	4 (bases 1 to 154746)	
REFERENCE	Barnett,B.C., Dolan,A., Telford,E.A., Davidson,A.J. and McGeoch,D.J.	A novel herpes simplex virus gene (UL49A) encodes a putative membrane protein with counterparts in other herpesviruses
JOURNAL	J. Gen. Virol. 73 (Pt 8), 2167-2171 (1992)	
MEDLINE	92356101	
PUBMED	1322965	
REFERENCE	5	
AUTHORS	Dolan,A.	Direct Submission
TITLE	Submitted (25-FEB-1997) A. Dolan, MRC Virology Unit, Church Street, Glasgow, G11 5UR, UK	
JOURNAL	revised by [6]	
REMARK	6 (bases 1 to 154746)	
REFERENCE	Dolan,A.	Direct Submission
AUTHORS	Submitted (08-DEC-1999) A. Dolan, MRC Virology Unit, Church Street, Glasgow, G11 5UR, UK	
TITLE	On Dec 13, 1999 this sequence version replaced gi:1869820.	
COMMENT	Location/Qualifiers	
FEATURES	1. 154746	/organism="Human herpesvirus 2"
source	/mol_type="genomic DNA"	
	/db_xref="taxon:10310"	
	1. .254	
	/note="a sequence"	
	255	
	/note="start of TRL (terminal copy of Long Repeat region)"	
	440. .1379	
	/gene="RL1"	
	join(440. .934,1089. .1379)	
	/gene="RL1"	
	/codon_start=1	
	/protein_id="CAB06759.1"	
	/db_xref="GI:1869821"	
	/db_xref="SWISS-PROT:P28283"	
	/translation="MSRRPGRRRGPPRRPRGAPVPRGAPVPPGALPTADSQMVPADSGTAVASPAASSLIRMLVQADSDADYAGNDABMANSPPSEGGKAPKAPPAARAAACPPPRKGGKGPQRLPHALRTTETVLATSLRRRPPASPADAPKACVCFSPRVQVRLVAVETARLARLRSSMRERADRRFRRVAAAVIGPCLEPEKARARARARHEDGPAEEEAARSSAAGPGRRAV"	
	440. .934	
	/gene="RL1"	
	/number=1	
	446. .460	
	/rpt_family="7"	
	461. .547	
	/rpt_family="6"	
	943. .1070	
	/rpt_family="5"	
	1089. .1379	
	/gene="RL1"	
	/product="neurovirulence factor"	
	/number=2	
	1738. .1743	
	2303. .5368	
	/gene="RL2"	
	join(2303. .2377,2785. .3462,3644. .5368)	

```

/gene="RL2"
/codon_start=1
/protein_id="CAB06760.1"
/db_xref="GI:1869822"
/db_xref="SWISS-PROT:p28284"
/translation="MEPRGTSRADPEPERPPROTPTGTPAAPHAMGLINDMOMLAS
SDSEETETGISDDDLHRDSTSEAGSTTEMPEKGLMDAAPAPPAEREGSGPTPAD
AQSGGGGGGVEEBAAGGGGVCVCTDETLAPLRQCSFPCLHPFCPCMKMTIPLIR
MTPCPLCPVALILYGVATYASSGSFSTIPVNDPRTRVEBAAVRAGTAVDPITWNPPT
APRSLSLGHTVRLALSTPTPMPGTDEDDDLADVDVYPAAPRAAROGGAGATRGIG
SQAPATRAPAPCAPSSSSGGAPLAAGVSGGGGPAVAAVPVAISLPPAAGGRAG
ARARVEDAAAEAGRTPPAPROPAAQEPVIVSDPSPRRRACGPLSFSSSSAOW
SSPGCGGGLPGSSGGAAPRAAVAPRVSPRAAAPVSAADACGAPAVYVADL
RAPERMTAOCTDPTOAGSLGRAGATDARGSGCGPAGCPGCTGTGCAAPVAABEA
AARPRKRGSDSGPAASSSSASSAAPPRLAPQGTGARRAPRAPDSGDRGHGPE
APASGAAPPSKSPSSQAALVAASSSSASSSSASSSSASSSSASSSSASSSSA
SSSAGACGASVASAGAGERRETSLGPRAAAPRGRCARTRHAEGGPEGADPAP
GLTRILPLTAGVSSVALAPYVNTVGTGDLPLDMETSHIGAYVVLVDGTNVADLIR
AAAPMSRRTLPEHARACVPRPDVPTPPASBWSLWMTPIGNMLPDGTLVGLADFTH
GURSHPSREGCAPAACDAPAGHE"
exon
2303..2377
/gene="RL2"
/number=1
/2785..3462
/gene="RL2"
/number=2
3468..3607
/tpc_family="4"
3644..5368
/gene="RL2"
/number=3
/functions="modulator of cell state and gene expression"
4655..4771
/tpc_family="3"
complement(4811)
/notes="proposed LAT splice acceptor site"
5618..5623
complement(7025)
/notes="proposed LAT splice donor site (5' end of stable
LAT species)"
complement(7732)
/notes="primary transcription initiation site for latency
associated transcripts (LAT's)"
9047..9193
/tpc_family="2"
9276..9297
/tpc_family="1"
9298
/notes="start of UL (long Unique region)"
9427..10101
/gene="UL1"
9427..10101
/gene="UL1"
/codon_start=1
/product="viral glycoprotein L"
/protein_id="CAB06761.1"
/db_xref="GI:1869823"
/db_xref="SWISS-PROT:p28278"
/translation="MGPTCLRGIVYMGMGAGMSQATETVLRSVIAKEVDILRVCC
MRTPADVMSRIRBAPSVIDYAIIDGIFLRHCPGIDTILMRHQRATLVNPLPAAAG
PLEDLSHVFPADTGETTRRALYKIRIDALGSRQAVSAHPVAGCGVYSTRKCG
VGRDRIARANTTSTWEPVSSDDEASQSKPLATPVTALASNAPRRVRSVTRGRRRRH
TRLRPN"
gene
10211..10978
/gene="UL2"
10211..10978
/gene="UL2"
/codon_start=1
/product="uracil-DNA glycosylase"
/protein_id="CAB06762.1"
/db_xref="GI:1869824"
/db_xref="SWISS-PROT:p28275"
/translation="NFSASTYPEPLGISGQATPPLPTSVPLDNAAFFRAFLIDDAWR

```

[illegible]

LAIIYGRIEREHQISPGDFPNLKRMQDLQAQDFSKFQPLKSKLLLEVVDMLANDIAQ  
LMLVROEESQRPIDQWVGGAFFGTLHGPFQHGYGEGAGESIDDAEMVAVARDKPMYDS  
IFYTLSPYDQKITGANAKKENVRSKLPSNVLGKIWKLADIDKDGMLDDDEFALANHLI  
KVXLEGHELPHNELPAHILLPRPRGKLPSDGVGSHSDGC"

BASE COUNT 541 a 639 c 653 g 448 t  
ORIGIN

Query Match 67.9%; Score 22.4; DB 9; Length 2281;  
Best Local Similarity 88.5%; Pred. No. 1.1e+02;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 4 TGCAACTCGTGTCTCCAGCATGC 29  
|||||  
|

Db 1550 TGCAACTCGTGTCTCCAGCATGC 1525  
|||||

Search completed: September 16, 2003, 19:14:19  
Job time : 724.713 secs



```

PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
PI Esser KM, Leary JJ;
XX
XX MPI, 1998-286847/25.
DR
DR P-PSDB; AAM72113.
XX
XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
XX
XX Claim 1; Page 406-407; 748pp; English.
XX
XX This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC sequence of the invention. This sequence was isolated from HSV-2 strain
CC SBS (deposited as ATCC VR-2546), is designated Contig ID 14, and encodes
CC a HSV-2 protein. The protein can be used for the treatment or
CC prevention of disease, to induce an immunological response in a mammal or
CC to identify inhibitors, activators or novel antivirals. Antagonists of
CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC or a vector containing it can also be used to induce an immunological
CC response in a mammal.
XX
XX Sequence 2645 BP; 404 A; 837 C; 892 G; 512 T; 0 other;
SQ
Query Match 68.5%; Score 22.6; DB 19; Length 2645;
Best Local Similarity 95.7%; Pred. No. 17;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 6 CAAACTCGTGTCTCTCCAGCATG 28
DB 755 CAAACTCGTGTCTCTCCAGCATG 777

RESULT 2
ID AAN71399/C
XX AAN71399 standard; DNA; 2724 BP.
XX
XX AAN71399;
AC
XX 25-MAR-2003 (updated)
DT 30-APR-1991 (first entry)
XX
XX Herpes Simplex Virus-2 gB from p52BXX.
DE
XX
XX Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;
KW glycoprotein; gB; ss.
XX
XX Herpes simplex virus type 2 HGS2.
OS
XX
XX US4642333-A.
PN
XX
XX 10-FEB-1987.
PD
XX
XX 20-JUN-1984; 84US-0622496.
PE
XX
XX 20-JUN-1984; 84US-0622496.
PR
XX
XX (PERS/) PERSON S.
PA
XX
XX Person S;
PI
XX
XX MPI, 1987-056354/08.
DR
DR P-PSDB; AAM71136.
XX
XX Amino acid chain of glyco:protein B of HSV-1 and 2 - prepd. as
PT recombinant and used for vaccines for herpes simplex virus types 1
PT and 2.
XX
XX Example; Table 2; 16pp; English.
XX
XX 'N' indicated nucleotide to be determined (corresp. to residues
CC 17-41 of HSV-1 gB).
CC
XX A pure non-glycosylated amino acid (AA) chain comprising a sequence
PC

```

```

CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2
CC virus which is antigenic to HSV-1 of HSV-2, which contains no more
CC than 750 AA residues, and which includes AA residues 135-649
CC inclusive is claimed. It can be used to produce vaccines for
CC prophylaxis and treatment of HSV-1 and HSV-2.
CC See AAN71303 for the HSV-1 sequence.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
XX Sequence 2724 BP; 494 A; 911 C; 854 G; 390 T; 75 other;
SQ
Query Match 68.5%; Score 22.6; DB 8; Length 2724;
Best Local Similarity 95.7%; Pred. No. 18;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 6 CAAACTCGTGTCTCTCCAGCATG 28
DB 2050 CAAACTCGTGTCTCTCCAGCATG 2028

RESULT 3
ID AAN50517/C
XX AAN50517 standard; DNA; 2727 BP.
XX
XX AAN50517;
AC
XX 21-NOV-1991 (first entry)
DT
XX
XX Sequence encoding the glycoprotein B (gB) of Herpes simplex virus
DE type 2 (HSV-2).
DE
XX
XX Vaccine; glycoprotein B; ss.
KW
XX
XX Herpes simplex virus type 2;
OS
XX
XX Key Location/Qualifiers
FH CDS 1..2727
FT CDS /*tag= a
FT
XX
XX EPI33063-A.
PN
XX
XX 13-FEB-1985.
PD
XX
XX 22-JUN-1984; 84EP-0401312.
PE
XX
XX 16-SEP-1983; 83US-0532996.
PR 23-JUN-1983; 83US-0506986.
PR 20-JUN-1984; 84US-0622496.
XX
XX (PERS/) PERSON S.
PA
XX
XX Person S;
PI
XX
XX MPI, 1985-039636/07.
DR
XX
XX New amino acid chains of glyco-protein B of herpes simplex virus
PT - prepd. by recombination DNA methods for use in vaccines
PT
XX
XX Disclosure; Table 2, Page 35-35c; 40pp; English.
XX
XX The inventors claim a non-glycosylated amino acid chain comprising a
CC sequence corresponding to that occurring in glycoprotein B of HSV-1
CC or HSV-2. Preferably, when the chain comprises a sequence corresp.
CC to that occurring in glycoprotein B of HSV-2, it contains 4-750 AA
CC residues. The chain esp. has a molecular wt. of 65000 daltons and
CC includes AA residues 155-629 from the HSV-2 glycoprotein B sequence.
XX
XX Sequence 2727 BP; 493 A; 913 C; 855 G; 391 T; 75 other;
SQ
Query Match 68.5%; Score 22.6; DB 6; Length 2727;
Best Local Similarity 95.7%; Pred. No. 18;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 6 CAAACTCGTGTCTCTCCAGCATG 28

```



Db 2053 CAAACTGCTGCTCCTCCAGCATG 2031

## RESULT 4

AAV62146/c  
ID AAV62146 standard; DNA; 2841 BP.

AAV62146;

07-DEC-1998 (first entry)

HSV-2 strain SBS Contig ID 91 DNA sequence.

HSV-2 strain SBS; immunological response induction; therapy;  
KM antiviral identification; viral protein inhibitor; ss.

Herpes simplex virus type 2.

Key Location/Qualifiers  
CDS 358..2745  
/\*tag= a

MO9820016-A1.

14-MAY-1998.

31-OCT-1997; 97WO-US20016.

09-JUN-1997; 97US-0049018.

04-NOV-1996; 96US-0030279.

(SMIK ) SMITHKLINE BEECHAM CORP.

Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;  
Besser KM, Leary JJ;

WPI; 1998-286847/25.

P-PSDB; AAW72062.

Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
and treatment of infection or inducing immunological response in  
mammal

Claim 1; Page 284-285; 748bp; English.

This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA  
sequence of the invention. This sequence was isolated from HSV-2 strain  
CC-SBS (deposited as ATCC VR-2546), is designated Contig ID 91, and encodes  
a HSV-2 protein. The proteins can be used for the treatment or  
prevention of disease, to induce an immunological response in a mammal or  
to identify inhibitors, activators or novel antivirals. Antagonists of  
the proteins can be used to inhibit a viral polypeptide. The DNA sequence  
or a vector containing it can also be used to induce an immunological  
response in a mammal.

Sequence 2841 BP; 577 A; 921 C; 915 G; 425 T; 3 other;

Query Match 68.5%; Score 22.6; DB 19; Length 2841;  
Best Local Similarity 95.7%; Pred. No. 18;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

6 CAAACTGCTGCTCCTCCAGCATG 28

2071 CAAACTGCTGCTCCTCCAGCATG 2049

## RESULT 5

AAAT93650/c  
ID AAT93650 standard; DNA; 3471 BP.

AAAT93650;

26-FEB-1998 (first entry)

Herpes simplex virus type 1 glycoprotein B encoding DNA.

Herpes simplex virus; HSV; type I; type II; glycoprotein B; vaccine;  
membrane protein; virus-specific glycoprotein;  
transmembrane anchor region; ss.

Herpes simplex virus type 1.

Key Location/Qualifiers  
CDS 55..62  
/\*tag= a  
/note= "putative"

TATA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

polyA\_signal

US648079-A.

15-JUL-1997.

08-DEC-1994; 94US-0351875.

20-OCT-1986; 86US-0921730.

06-APR-1984; 84US-0597784.

20-SEP-1990; 90US-0587179.

21-DEC-1992; 92US-0993415.

18-OCT-1993; 93US-0138717.

(CHIR ) CHIRON CORP.

Burke RL, Pachl C, Valenzuela PDT;

WPI; 1997-372022/34.

P-PSDB; AAW34552.

Vaccines against herpes simplex virus infection - containing  
recombinant HSV glycoprotein B

Disclosure; Fig 4A-F; 33pp; English.

The present sequence is from the Herpes simplex virus (HSV) type I  
strain Patton and encodes a glycoprotein B. HSV is a double stranded  
virus packaged within an icosahedral nucleocapsid enveloped within a  
membrane. The membrane includes a number of virus-specific  
glycoproteins, with glycoprotein B being one of the most abundant.  
Glycoprotein B from both HSV type I and type II are cross reactive. New  
vaccines comprising recombinantly produced glycosylated glycoprotein B  
that has a deletion of at least part of the transmembrane anchor region,  
in combination with a carrier and an adjuvant have been produced. The  
vaccines are used for immunising humans against HSV (HSV type 1 or  
HSV type 2) by vaccination before or after primary infection with HSV.

Sequence 3471 BP; 639 A; 1201 C; 1069 G; 562 T; 0 other;

Query Match 68.5%; Score 22.6; DB 18; Length 3471;  
Best Local Similarity 95.7%; Pred. No. 18;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

6 CAAACTGCTGCTCCTCCAGCATG 28

2349 CAAACTGCTGCTCCTCCAGCATG 2327

```

RESULT 6
ID AAN80907/c      standard; DNA; 3472 BP.
XX
AC AAN80907;
XX
DT 25-MAR-2003 (updated)
DT 09-MAR-1992 (first entry)
XX
DE Sequence of Herpes simplex virus (HSV) glycoprotein B (gB) 1 DNA.
XX
KM Vaccine; herpes simplex virus; therapy; ss.
XX
OS Herpes simplex virus.
XX
FH Key      Location/Qualifiers
FT CDS      309..3023
FT          /*tag= a
XX
PA MO8802634-A.
XX
PI 21-APR-1988.
XX
PF 20-OCT-1987; 87WO-US02709.
XX
PR 20-JUL-1987; 87US-0079605.
PR 20-OCT-1986; 86US-0921213.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Burke LR, Pachl C, Valenzuela P;
XX
DR WPI; 1988-119368/17.
DR P-PSDB; AAP80914.
XX
PT Vaccine for treatment of herpes simplex virus - contains
PT recombinant HSV glyco:proteins B and D
XX
PS Example; Fig 4; 71pp; English.
XX
CC Prep. of recombinant gB and gD is described in WO88504587. The
CC amino acid sequence and DNA sequence for gB 1 presented in AAP80914
CC and AAN80907 differ from that originally presented in Table 1 of
CC International Publication No. WO 85/04587. The DNA sequence in Table
CC 1 contains an error in that an additional nucleotide (G) is listed
CC at position 607 which resulted in a shift in reading frame relative
CC to AAN80907 from which this nucleotide has been deleted.
CC (Updated on 25-MAR-2003 to correct PA field.)
CC
SQ Sequence 3472 BP; 639 A; 1202 C; 1068 G; 563 T; 0 other;

Query Match      68.5%; Score 22.6; DB 9; Length 3472;
Best Local Similarity 95.7%; Pred. No. 18;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAACTCGTGKTCCTCCAGCATG 28
DB 2349 CAAACTCGTGCTCTCCAGCATG 2327

RESULT 7
AAQ48496/c
ID AAQ48496 standard; DNA; 3472 BP.
XX
AC AAQ48496;
XX
DT 25-MAR-2003 (updated)
DT 29-MAR-1994 (first entry)
XX
DE Glycoprotein B (gB1).
XX
KM Glycoprotein; gB1; gB2; herpes simplex virus; HSV-1; HSV-2;
KM flanking; initiation; termination; transcription; translation;

```

```

KM regulatory sequence; ss.
XX
OS Herpes simplex virus 1 strain Patton.
XX
FH Key      Location/Qualifiers
FT CDS      309..3023
FT          /*tag= a
XX
PA US5244792-A.
XX
PI 14-SEP-1993.
XX
PF 20-SEP-1990; 90US-0587179.
XX
PR 06-APR-1984; 84US-0597784.
PR 20-OCT-1986; 86US-0921730.
PR 20-SEP-1990; 90US-0587179.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Burke RL, Pachl C, Valenzuela PDT;
XX
DR WPI; 1993-302641/38.
DR P-PSDB; AAR41778.
XX
PT DNA construct for expressing HSV glycosylated polypeptide -
PT useful for vaccinating against HSV-1 and -2 infections
XX
PS Disclosure; Fig 4; 33pp; English.
XX
CC DNA constructs for expressing a glycosylated polypeptide in a
CC eukaryotic host cell comprises (1) an oligonucleotide sequence (OS)
CC free of natural flanking sequences, encoding glycoprotein B (gB)
CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating
CC transcriptional and translational regulatory sequences flanking OS,
CC at least one of these sequences not being from HSV. The HSV-1 and
CC HSV-2 sequences are given in AAQ48496 and AAQ48497 respectively.
CC (Updated on 25-MAR-2003 to correct PF field.)
CC
SQ Sequence 3472 BP; 639 A; 1202 C; 1069 G; 562 T; 0 other;

Query Match      68.5%; Score 22.6; DB 14; Length 3472;
Best Local Similarity 95.7%; Pred. No. 18;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAACTCGTGKTCCTCCAGCATG 28
DB 2349 CAAACTCGTGCTCTCCAGCATG 2327

RESULT 8
AAV62176
ID AAV62176 standard; DNA; 117213 BP.
XX
AC AAV62176;
XX
DT 13-JAN-1999 (first entry)
XX
DE HSV-2 strain SBS Contig ID 15 DNA sequence.
XX
KM HSV-2 strain SBS; immunological response induction; therapy;
KM antiviral identification; viral protein inhibitor; ss.
XX
OS Herpes simplex virus type 2.
XX
FH Key      Location/Qualifiers
FT CDS      755..1297
FT          /*tag= a
FT          /product= "ORF#1 protein"
FT          /note= "encoded protein shown in AAW72170"
FT          1170..2174
FT          /*tag= b
FT          /product= "ORF#2 protein"

```

```

FT /note= "encoded protein shown in AAW72171"
FT CDS 2229..2930
FT /*tag= c
FT /product= "ORF#3 protein"
FT /note= "encoded protein shown in AAW72172"
FT CDS complement (3130..3735)
FT /*tag= d
FT /product= "ORF#4 protein"
FT /note= "encoded protein shown in AAW72173"
FT CDS complement (3802..6447)
FT /*tag= e
FT /product= "ORF#5 protein"
FT /note= "encoded protein shown in AAW72174"
FT CDS 6017..8482
FT /*tag= f
FT /product= "ORF#6 protein"
FT /note= "encoded protein shown in AAW72250"
FT CDS 6026..8482
FT /*tag= g
FT /product= "ORF#6f protein"
FT /note= "encoded protein shown in AAW72249"
FT CDS 6065..8482
FT /*tag= h
FT /product= "ORF#6e protein"
FT /note= "encoded protein shown in AAW72248"
FT CDS 6167..8482
FT /*tag= i
FT /product= "ORF#6d protein"
FT /note= "encoded protein shown in AAW72247"
FT CDS 6236..8482
FT /*tag= j
FT /product= "ORF#6c protein"
FT /note= "encoded protein shown in AAW72246"
FT CDS 6326..8482
FT /*tag= k
FT /product= "ORF#6b protein"
FT /note= "encoded protein shown in AAW72245"
FT CDS 6446..8482
FT /*tag= l
FT /product= "ORF#6a protein"
FT /note= "encoded protein shown in AAW72244"
FT CDS 8457..9347
FT /*tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in AAW72175"
FT CDS complement (9604..11855)
FT /*tag= n
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in AAW72176"
FT CDS complement (11905..14508)
FT /*tag= o
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in AAW72222"
FT CDS complement (11905..14520)
FT /*tag= p
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in AAW72223"
FT CDS 14399..15802
FT /*tag= q
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in AAW72177"
FT CDS complement (15996..16286)
FT /*tag= r
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in AAW72178"
FT CDS complement (16202..18064)
FT /*tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in AAW72179"
FT CDS complement (18105..19661)

FT /*tag= t
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in AAW72180"
FT CDS complement (19415..20074)
FT /*tag= u
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in AAW72181"
FT CDS 20155..21453
FT /*tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in AAW72182"
FT CDS complement (21326..22291)
FT /*tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in AAW72183"
FT CDS complement (22546..24654)
FT /*tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in AAW72184"
FT CDS 24684..25955
FT /*tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in AAW72185"
FT CDS complement (26295..27251)
FT /*tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in AAW72186"
FT CDS complement (27630..31754)
FT /*tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in AAW72225"
FT CDS complement (27630..31784)
FT /*tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in AAW72224"
FT CDS complement (32067..32735)
FT /*tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in AAW72187"
FT CDS 33140..34984
FT /*tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in AAW72226"
FT CDS 33386..34984
FT /*tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in AAW72227"
FT CDS complement (35205..37721)
FT /*tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in AAW72188"
FT CDS complement (38058..39188)
FT /*tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in AAW72189"
FT CDS 39090..39935
FT /*tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in AAW72190"
FT CDS 40216..41973
FT /*tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in AAW72191"
FT CDS 42206..44178
FT /*tag= aj
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in AAW72192"
FT CDS complement (44853..47297)
FT /*tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in AAW72193"
FT CDS 47122..47338

```

```

FT      /*tag= a1
FT      /product= "ORF#29 protein"
FT      /note= "encoded protein shown in AAW72194"
FT      /tag= am
FT      /product= "ORF#30 protein"
FT      /note= "encoded protein shown in AAW72195"
FT      /tag= an
FT      /product= "ORF#31 protein"
FT      /note= "encoded protein shown in AAW72196"
FT      /tag= ao
FT      /product= "ORF#32 protein"
FT      /note= "encoded protein shown in AAW72197"
FT      /tag= ap
FT      /product= "ORF#33 protein"
FT      /note= "encoded protein shown in AAW72198"
FT      /tag= aq
FT      /product= "ORF#34 protein"
FT      /note= "encoded protein shown in AAW72199"
FT      /tag= ar
FT      /product= "ORF#35 protein"
FT      /note= "encoded protein shown in AAW72200"
FT      /tag= as
FT      /product= "ORF#36 protein"
FT      /note= "encoded protein shown in AAW72201"
FT      /tag= at

```

```

Query Match      68.5%; Score 22.6; DB 19; Length 117213;
Best Local Similarity 95.7%; Pred. No. 29;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      6 CAAACTCGTGTCTCCAGCATG 28
DB      45257 CAAACTCGTGTCTCCAGCATG 45279

RESULT 9
AID25519
ID      AAD25519 standard; DNA; 154746 BP.
AC      AAD25519;
XX      26-MAR-2002 (first entry)
DE      Human herpesvirus 2 complete DNA genome.
XX      Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
XX      antibacterial; fungicide; protozoicide; antitubercular; antiinflammatory;
XX      antitachytic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
XX      immune response; vasculitic; vaccine; gene therapy; autoimmune disease;
XX      vasculitis; de.
XX      Human herpesvirus 2.
XX      WO200176643-A1.
XX      18-OCT-2001.
XX      06-APR-2001; 2001WO-US11372.
XX      07-APR-2000; 2000US-195680P.
XX      (BAVU ) BAYLOR COLLEGE MEDICINE.
XX      Orson FM, Kinney BM, Bhogal BS;
XX

```

```

DR      WPI; 2002-066308/09.
XX      Composition for oral delivery of vaccines, comprises expression vector
XX      containing antigenic genomic sequence, bound to aggregated
XX      protein-polycationic polymer conjugate or suspension
XX      Disclosure; Page 90-132; 145pp; English.
XX      The invention relates to a composition comprising an expression vector
XX      bound to an aggregated protein-polycationic polymer conjugate or
XX      suspension. The expression vector contains a promoter polynucleotide
XX      sequence operatively linked to a polynucleotide sequence encoding an
XX      antigen which is a fragment of a gene or genome associated with an
XX      infectious disease, cancer and autoimmune disease such as rheumatoid
XX      arthritis, vasculitis, and multiple sclerosis, pathogenic genes
XX      consisting of bacterium, fungus, protozoa and virus such as human
XX      immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX      virus (HCV), influenza and respiratory syncytial virus (RSV), and
XX      optionally comprising a nucleotide sequence encoding a cytokine (or a
XX      cytokine expression vector), is useful for inducing an immune response
XX      (systemic and/or mucosal) in an organism. The cytokine expression vector
XX      contains a sequence for granulocyte macrophage-colony stimulating factor
XX      (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
XX      the antigen and the cytokine are under transcriptional control of same or
XX      different promoter polynucleotide sequences. The expression vector, as a
XX      DNA vaccine is useful for treating a condition in an organism. The
XX      present sequence is human herpesvirus 2 complete DNA genome related
XX      to the invention.
XX      Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 other;

```

```

Query Match      68.5%; Score 22.6; DB 24; Length 154746;
Best Local Similarity 95.7%; Pred. No. 30;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      6 CAAACTCGTGTCTCCAGCATG 28
DB      54077 CAAACTCGTGTCTCCAGCATG 54099

RESULT 10
AID12477c
ID      AAI61247 standard; CDNA; 2550 BP.
AC      AAI61247;
XX      22-OCT-2001 (first entry)
DE      Human polynucleotide SEQ ID NO 5236.
XX      Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX      peripheral nervous system; neuropathy; central nervous system; CNS;
XX      Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX      amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX      chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX      leukaemia; ss.
XX      Homo sapiens.
XX      WO200153312-A1.
XX      26-JUL-2001.
XX      26-DEC-2000; 2000WO-US34263.
XX      21-JAN-2000; 2000US-0488725.
XX      25-APR-2000; 2000US-0552317.
XX      09-JUL-2000; 2000US-0558042.
XX      19-JUL-2000; 2000US-0620312.
XX      03-AUG-2000; 2000US-0653450.
XX      14-SEP-2000; 2000US-0662191.
XX      19-OCT-2000; 2000US-0693036.
XX      29-NOV-2000; 2000US-0727344.

```

XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Auendi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QH, Zhou P, Goodrich R, Drmanac RT;  
 XX WPI; 2001-442253/47.  
 DR P-PSDB; AAM42091.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Claim 1; SEQ ID NO 5236; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilization of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 2550 BP; 633 A; 686 C; 650 G; 581 T; 0 other;  
 XX  
 Query Match 67.9%; Score 22.4; DB 22; Length 2550;  
 Best Local Similarity 88.5%; Pred. No. 21;  
 Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 TGCAAACTCGTGTCTCTCCAGCATGC 29  
 Db 754 TGCAAACTCGTGTCTCTCCAGCATGC 729

RESULT 11  
 AAL53531/c  
 ID AAL53531 standard; cDNA; 2815 BP.  
 XX  
 AC AAL53531;  
 XX  
 DT 30-JAN-2003 (first entry)  
 XX  
 DE cDNA of human EHD2 gene, SEQ ID NO 7.  
 XX  
 KW Cytostatic; osteopathic; recombinant Eps15 homology; EH domain; EHD1;  
 KW EHD2; endocytosis; IGf1 signaling; suppressing adhesion; invasion;  
 KW metastasis; bone formation; osteoporosis; human; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 51..1001  
 FT /\*tag= a  
 FT /product= "Human EHD2 protein"  
 XX  
 PN US2002115069-A1.  
 XX  
 PD 22-AUG-2002.  
 XX  
 PF 17-MAY-1999; 99US-0312762.  
 XX  
 PR 21-FEB-1997; 97IL-0120283.  
 PR 20-FEB-1998; 98US-0026898.  
 XX

PA (HORO/) HOROWITZ M.  
 PA (MINT/) MINTZ L.  
 XX  
 PI Horowitz M, Mintz L;  
 XX  
 DR WPI; 2002-731288/79.  
 DR P-PSDB; AAO26366.  
 XX  
 XX Novel recombinant Eps15 homology domain containing protein useful for  
 PT regulating endogenous EHD protein by regulating insulin-like growth  
 PT factor 1 receptor cell signaling via altered clathrin coated pit  
 PT mediated endocytosis -  
 XX  
 PS Claim 6; page 45-46; 88bp; English.  
 XX  
 CC The invention relates to a novel recombinant Eps15 homology (EH) domain  
 CC containing protein 1 or 2 (EHD1 or EHD2) with an N-terminal region  
 CC containing a nucleotide binding consensus site, a central coiled coil  
 CC structure and a C-terminal region including an EH domain, where the  
 CC polypeptide participates in endocytosis. The peptides, peptide analogues  
 CC and/or the polynucleotide sequences are useful for regulating EHD1/2  
 CC protein activity in vivo, where the EHD1/2 protein and its gene are  
 CC useful for upregulating the protein activity. The novel method is useful  
 CC for lowering the rate of IGf1 signaling and therefore useful for  
 CC suppressing adhesion, invasion and metastasis where EHD is overexpressed,  
 CC and therefore endocytosis. The method is also useful for elongating IGf1  
 CC effects and therefore for increasing bone formation in osteoporosis  
 CC conditions. This polynucleotide sequence represents the cDNA of a human  
 CC EHD2 gene of the invention.  
 XX  
 SQ Sequence 2815 BP; 695 A; 754 C; 730 G; 636 T; 0 other;  
 XX  
 Query Match 67.9%; Score 22.4; DB 24; Length 2815;  
 Best Local Similarity 88.5%; Pred. No. 21;  
 Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 TGCAAACTCGTGTCTCTCCAGCATGC 29  
 Db 980 TGCAAACTCGTGTCTCTCCAGCATGC 955

RESULT 12  
 AAI59461/c  
 ID AAI59461 standard; cDNA; 3327 BP.  
 XX  
 AC AAI59461;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 1664.  
 XX  
 KW Human; neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 51..1001  
 FT /\*tag= a  
 FT /product= "Human EHD2 protein"  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 XX

```

PR    29-NOV-2000; 2000US-0727344.
XX
XX    (HYSE-) HYSEQ INC.
PA
P1    Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
P1    Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
P1    Zhao QH, Zhou P, Goodrich R, Drmanac RT;
XX
XX    WPI, 2001--442253/47.
DR
DR    P-P8DB; AAM40305.
XX
XX    Novel nucleic acids and polypeptides, useful for treating disorders
PT    such as central nervous system injuries -
XX
XX    Claim 1; SEQ ID NO 1664; 10078bp; English.
XX
XX    The invention relates to human nucleic acids (AA157798-AA161369) and
CC    the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
CC    immunosuppressant and cytoskeletal activity. The polynucleotides are useful
CC    in gene therapy. A composition containing a polypeptide or polynucleotide
CC    of the invention may be used to treat diseases of the peripheral nervous
CC    system, such as peripheral nervous injuries, peripheral neuropathy and
CC    localised neuropathies and central nervous system diseases, such as
CC    Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC    lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC    utilisation of the activities such as: Immune system suppression,
CC    Activation/inhibit activity, chemotactic/chemokinetic activity, haemostatic
CC    and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC    assays for receptor activity, arthritis and inflammation, leukaemia and
CC    C.N.S disorders.
CC
CC    Note: The sequence data for this patent did not form part of the printed
CC    specification.
XX
SQ    Sequence 3327 BP; 814 A; 900 C; 880 G; 733 T; 0 other;
XX
XX    Query Match          67.9%; Score 22.4; DB 22; Length 3327;
XX    Best Local Similarity 88.5%; Pred. No. 22;
XX    Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0
OY      4 TGCAACTCGTGCTCCTCCAGCATGC 29
DB      1522 TGCAACTCGTGTGCTCCTCCAGCATGC 1497
        |||||
RESULT 13
ABZ41610/C
ID      ABZ41610 standard; DNA; 1494 BP.
XX
XX    ABZ41610;
AC
XX
XX    07-MAR-2003 (first entry)
DT
XX
XX    N. gonorrhoeae nucleotide sequence SEQ ID 7809.
DE
XX
XX    Antibacterial; infection; vaccine; gene therapy; gene; de.
KW
XX
XX    Neisseria gonorrhoeae.
OS
XX
XX    WO200279243-A2.
PN
PD      10-OCT-2002.
XX
PP      12-FEB-2002; 2002WO-IB02069.
PR
PR      12-FEB-2001; 2001GB-0003424.
XX
PA      (CHIR-) CHIRON SPA.
XX
PI      Fontana MR, Pizza M, Maignani V, Monaci E;
XX    WPI, 2003-058415/05.
XX    P-P8DB; ABP80640.
XX

```

```

PT  New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT  medicament for treating or preventing N. gonorrhoeae infection
PS  Disclosure; Page 759; 815pp; English.
XX
XX  The present invention relates to proteins from Neisseria gonorrhoeae.
CC  Also disclosed are the nucleic acid molecules encoding the proteins and
CC  antibodies that specifically bind to the proteins. The composition
CC  comprising the protein, nucleic acid or antibody is useful for the
CC  manufacture of a medicament for treating or preventing N. gonorrhoeae
CC  infection. This may be in the form of a vaccine or gene therapy.
CC  Sequences given in records AB237706-AB242016 represent nucleic acid
CC  molecules of the invention.
XX
SQ  Sequence 1494 BP; 304 A; 453 C; 432 G; 305 T; 0 other;
Query Match 64.8%; Score 21.4; DB 25; Length 1494;
Best Local Similarity 75.8%; Pred. No. 50;
Matches 25; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
OY 1 CCCGCAACCTCGTGTCTCTCAGCAGCAGG 33
Db 324 CCCGCAACCTCGTGTCTCTCAGCAGCAGG 292
RESULT 14
AA014479/C
ID AA014479 standard; DNA; 2088 BP.
XX
XX  AA014479;
AC 25-MAR-2003 (updated)
XX
DT 23-JUN-1992 (first entry)
XX
DE Truncated HSVGB gene.
XX
XX  Vaccine; antigen; ss.
XX
XX  Herpes simplex virus.
OS
FH Key Location/Qualifiers
FT CDS 1..2088
FT /tag= a
XX
XX  JP03218397-A.
XX
XX 25-SEP-1991.
XX
XX 21-JUN-1990; 90JP-0161448.
XX
XX 30-NOV-1989; 89JP-0308941.
XX
XX 22-JUN-1989; 89JP-0158238.
XX
XX 21-JUN-1990; 90JP-0161448.
XX
XX (TAKEDA CHEM IND LTD.
XX
XX MPI; 1991-326937/45.
XX
XX P-PSDB; AAR14666.
XX
XX HSVGB polypeptide(s) obtd. by recombinant DNA techniques -
PT useful as vaccines against HSV and in diagnosis, can be produced
PT cheaply and safely.
XX
XX Claim 3; Fig 7; 24pp; Japanese.
XX
XX The sequence encodes a truncated form of the HSVGB polypeptide.
CC The recombinant protein can be used to prepare vaccines for
CC prophylaxis of HSV infection and for use in diagnostic kits.
CC See also AA014478.
XX (updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 2088 BP; 417 A; 751 C; 620 G; 300 T; 0 other;

```

Query Match 64.8%; Score 21.4; DB 12; Length 2088;  
 Best Local Similarity 95.7%; Pred. No. 52;  
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CAAACTCGTGTCTCTCCGCGCATG 28  
 |||||  
 DB 1963 CAAACTCGTGTCTCTCCGCGCATG 1941

## RESULT 15

AAW71303/C  
 ID AAW71303 standard; DNA; 2712 BP.

XX AAW71303;

AC 25-MAR-2003 (updated)

DT 30-APR-1991 (first entry)

XX Herpes Simplex Virus-1 GB from pKEXX.

DE Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;

KM glycoprotein; gB, gS.

XX Herpes simplex virus type 1 (KOS).

OS US464233-A.

XX 10-FEB-1987.

PD 20-JUN-1984; 84US-0622496.

XX 20-JUN-1984; 84US-0622496.

XX (PERS/) PERSON S.

PA Person S;

PI WPI; 1987-056354/08.

XX P-PSDB; AAP71135.

DR Amino acid chain of glycoprotein B of HSV-1 and 2 - prepd. as

XX recombinant and used for vaccines for herpes simplex virus types 1

PT and 2.

XX Example; Table 2; 16pp; English.

PS A pure non-glycosylated amino acid (AA) chain comprising a sequence

XX corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2

CC virus which is antigenic to HSV-1 of HSV-2, which contains no more

CC than 750 AA residues, and which includes AA residues 135-649

CC inclusive is claimed. It can be used to produce vaccines for

CC prophylaxis and treatment of HSV-1 and HSV-2.

CC See AAW71399 for the HSV-2 sequence.

CC (Updated on 25-MAR-2003 to correct PR field.)

XX Sequence 2712 BP; 519 A; 945 C; 849 G; 397 T; 2 other;

XX

QY Query Match 64.8%; Score 21.4; DB 8; Length 2712;

XX Best Local Similarity 95.7%; Pred. No. 54;

CC Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX

DB 2047 CAAACTCGTGTCTCTCCGCGCATG 2025

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Search completed: September 16, 2003, 18:37:07  
 Job time : 167.327 secs

**BEST AVAILABLE COPY**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:24:35 ; Search time 42.1485 Seconds  
(without alignments)  
345.579 Million cell updates/sec

Title: US-09-594-065-4

Perfect score: 1 ccctgcaactcgtgkctccagcatgcagg 33

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	22.6	68.5	5244792-1	Patent No. 5244792
2	22.4	67.9	2815	Sequence 7, Appl1
3	21.4	64.8	2713	Sequence 13, Appl1
4	21.4	64.8	2713	Sequence 13, Appl1
5	21.4	64.8	3472	Patent No. 5244792
6	20.2	61.2	2517	Sequence 1, Appl1
7	20.2	61.2	2939	Sequence 3, Appl1
8	20.2	61.2	3138	Sequence 1, Appl1
9	19.4	58.8	4403765	Sequence 2, Appl1
10	19.4	58.8	4411529	Sequence 1, Appl1
11	19.2	58.2	1001	Sequence 358, App
12	19.2	58.2	1001	Sequence 358, App
13	18.8	57.0	1656	Sequence 27, Appl1
14	18.8	57.0	3955	Sequence 7, Appl1
15	18.8	57.0	4464	Sequence 7, Appl1
16	18.8	57.0	4483	Sequence 7, Appl1
17	18.6	56.4	1176	Sequence 42, Appl1
18	18.6	56.4	1887	Sequence 9, Appl1
19	18.6	56.4	2573	Sequence 609, App
20	18.6	56.4	1230025	Sequence 1, Appl1
21	18.4	55.8	2943	Sequence 4, Appl1
22	18.4	55.8	3177	Sequence 4, Appl1
23	18.2	55.2	3348	Sequence 2, Appl1
24	18.2	55.2	3635	Sequence 6, Appl1
25	18.2	55.2	7812	Sequence 1, Appl1
26	18.2	55.2	8453	Sequence 45, Appl1
27	18.2	55.2	14707	Sequence 3, Appl1

28	18	54.5	1112	4	US-09-663-600A-136	Sequence 136, App
29	17.8	53.9	2311	2	US-08-976-259-26	Sequence 26, Appl1
30	17.8	53.9	3627	2	US-08-232-087A-1	Sequence 1, Appl1
31	17.8	53.9	5084	1	US-08-306-691B-21	Sequence 21, Appl1
32	17.8	53.9	5084	5	PCT-US93-06251-25	Sequence 25, Appl1
33	17.6	53.3	624	4	US-09-221-017B-43	Sequence 43, Appl1
34	17.6	53.3	795	4	US-09-252-991A-3803	Sequence 3803, Ap
35	17.6	53.3	1122	3	US-08-746-883-1	Sequence 1, Appl1
36	17.6	53.3	1623	4	US-09-252-991A-3844	Sequence 3844, Ap
37	17.6	53.3	4167	4	US-09-252-991A-3666	Sequence 3666, Ap
38	17.6	53.3	8802	3	US-08-896-449A-1	Sequence 1, Appl1
39	17.6	53.3	8802	3	US-09-132-652-1	Sequence 1, Appl1
40	17.4	52.7	576	4	US-09-252-991A-13853	Sequence 13853, A
41	17.4	52.7	1098	4	US-09-252-991A-13595	Sequence 13595, A
42	17.4	52.7	1340	4	US-09-000-062-2	Sequence 2, Appl1
43	17.4	52.7	1340	4	US-09-000-062-4	Sequence 4, Appl1
44	17.4	52.7	1340	4	US-08-945-144A-2	Sequence 2, Appl1
45	17.4	52.7	1340	4	US-08-945-144A-4	Sequence 4, Appl1

## ALIGNMENTS

```
RESULT 1
5244792-1/c
Patent No. 5244792
APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.
TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYOPIROTEIN
FROM HERPES SIMPLEX VIRUS
NUMBER OF SEQUENCES: 19
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/587,179
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 921,730
FILING DATE: 20-OCT-1986
APPLICATION NUMBER: 597,784
FILING DATE: 06-APR-1984
SEQ ID NO:1:
5244792-1
LENGTH: 3472
Query Match 68.5%; Score 22.6; DB 6; Length 3472;
Best Local Similarity 95.7%; Pred. No. 1.6;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

6 CAAACTGCTGCTCTCCAGCATG 28  
Db 2349 CAAACTGCTGCTCTCCAGCATG 2327

RESULT 2  
US-09-312-762A-7/c  
Sequence 7, Application US/09312762A  
Patent No. 6552177  
GENERAL INFORMATION:  
APPLICANT: MIA HOROWITZ ET AL.  
TITLE OF INVENTION: RH DOMAIN CONTAINING GENES AND PROTEINS  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Mark M. Friedman c/o Anthony Caetorina  
STREET: 2001 Jefferson Davis Highway, Suite 207  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk  
COMPUTER: Twinhead\* Slimnote-890TX  
OPERATING SYSTEM: MS DOS version 6.2,  
OPERATING SYSTEM: Windows version 3.11  
SOFTWARE: word for Windows version 2.0 converted to  
an ASCII file

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/312,762A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/026,898  
FILING DATE: 20 FEB 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Friedman, Mark M.  
REGISTRATION NUMBER: 33,883  
REFERENCE/DOCKET NUMBER: 916/10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 972-3-5625553  
TELEFAX: 972-3-5625554  
TELEX:  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2815  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-312-762A-7

Query Match 67.9%; Score 22.4; DB 4; Length 2815;  
Best Local Similarity 88.5%; Pred. No. 1.9;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 TGCACCTGCTGCTCTCCAGCATG 29  
DB 980 TGCACCTGCTGCTCTCCAGCATG 955

RESULT 3  
US-08-804-439A-13/C  
Sequence 13, Application US/08804439A  
Patent No. 6015565  
GENERAL INFORMATION:  
APPLICANT: Rose, Timothy M.  
APPLICANT: Bosch, Marix L.  
APPLICANT: Strand, Kurt  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RRV/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Ste 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,439A  
FILING DATE: February 21, 1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Hall, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09176/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-804-439A-13  
Query Match 64.8%; Score 21.4; DB 3; Length 2713;  
Best Local Similarity 95.7%; Pred. No. 5;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CAACCTGCTGCTCTCCAGCATG 28  
DB 2047 CAACCTGCTGCTCTCCAGCATG 2025

RESULT 4  
US-08-720-229-13/C  
Sequence 13, Application US/08720229  
Patent No. 6022542  
GENERAL INFORMATION:  
APPLICANT: Rose, Timothy M.  
APPLICANT: Bosch, Marix L.  
APPLICANT: Strand, Kurt  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RRV/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 100  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/720,229  
FILING DATE: 26-SEP-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Schiff, J. Michael  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20002.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-720-229-13

Query Match 64.8%; Score 21.4; DB 3; Length 2713;  
Best Local Similarity 95.7%; Pred. No. 5;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CAACCTGCTGCTCTCCAGCATG 28  
DB 2047 CAACCTGCTGCTCTCCAGCATG 2025

RESULT 5  
5244792-2/C  
Patent No. 5244792  
APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.  
TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYCOPROTEIN  
B FROM HERPES SIMPLEX VIRUS  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/587,179  
FILING DATE: 20-SEP-1990

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 921,730  
FILING DATE: 20-OCT-1986  
APPLICATION NUMBER: 597,784  
FILING DATE: 06-APR-1984  
SEQ ID NO: 2:  
LENGTH: 3472  
5244792-2

Query Match 64.8%; Score 21.4; DB 6; Length 3472;  
Best Local Similarity 95.7%; Pred. No. 5.2;  
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CAAACTCGTGKTCCTCCAGCATG 28  
DB 2358 CAAACTCGTGATCCTCCAGCATG 2336

## RESULT 6

US-07-906-930E-1  
Sequence 1, Application US/07906930E  
Patent No. 5534631  
GENERAL INFORMATION:  
APPLICANT: Gaynor, Richard B.  
APPLICANT: Nirula, Ajay  
TITLE OF INVENTION: DNA ENCODING THE INTERLEUKIN BINDING  
TITLE OF INVENTION: FACTOR (ILF)  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
ZIP: 77210  
COUNTRY: USA  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/906,930E  
FILING DATE: 30-JUN-1992  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Serlich, Gary J.  
REGISTRATION NUMBER: 34,430  
REFERENCE/DOCKET NUMBER: UTSD:262/SER  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512-418-3000  
TELEFAX: 512-474-7577  
TELEX: NOT APPLICABLE  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2517 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-07-906-930E-1

Query Match 61.2%; Score 20.2; DB 1; Length 2517;  
Best Local Similarity 75.8%; Pred. No. 16;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCCTGCAAACTCGTGKTCCTCCAGCATGCAAGG 33  
DB 255 CGCTGCAAACTCTGCGCCCTCCAGCCCCCGGGG 287

## RESULT 7

## US-07-906-930E-3

Sequence 3, Application US/07906930E  
Patent No. 5534631  
GENERAL INFORMATION:  
APPLICANT: Gaynor, Richard B.  
APPLICANT: Nirula, Ajay  
TITLE OF INVENTION: DNA ENCODING THE INTERLEUKIN BINDING  
TITLE OF INVENTION: FACTOR (ILF)  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
ZIP: 77210  
COUNTRY: USA

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/906,930E  
FILING DATE: 30-JUN-1992  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Serlich, Gary J.  
REGISTRATION NUMBER: 34,430  
REFERENCE/DOCKET NUMBER: UTSD:262/SER  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512-418-3000  
TELEFAX: 512-474-7577  
TELEX: NOT APPLICABLE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2939 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "DNA"  
US-07-906-930E-3

Query Match 61.2%; Score 20.2; DB 1; Length 2939;  
Best Local Similarity 75.8%; Pred. No. 16;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCCTGCAAACTCGTGKTCCTCCAGCATGCAAGG 33  
DB 255 CGCTGCAAACTCTGCGCCCTCCAGCCCCCGGGG 287

## RESULT 8

US-09-434-408-1/C  
Sequence 1, Application US/09434408  
Patent No. 6440697  
GENERAL INFORMATION:  
APPLICANT: Venezia, Domenick  
APPLICANT: Grossmann, Angelika  
TITLE OF INVENTION: RING FINGER PROTEIN ZAPOR3  
FILE REFERENCE: 98-41  
CURRENT APPLICATION NUMBER: US/09/434,408  
CURRENT FILING DATE: 1999-11-04  
EARLIER APPLICATION NUMBER: US 60/108,258  
EARLIER FILING DATE: 1998-11-12  
NUMBER OF SEQ ID NOS: 23  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 3138  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:

/ NAME/KEY: CDS  
/ LOCATION: (367) ... (2535)  
US-09-434-408-1

Query Match 61.2%; Score 20.2; DB 4; Length 3138;  
Best Local Similarity 81.5%; Pred. No. 16;  
Matches 22; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCTGCAAACTCGTGTCTCCAGCAT 27  
Db 1102 CCTGCAAACTCGAGTTCTCCAGCAT 1156

RESULT 9  
US-09-103-840A-2/c  
/ Sequence 2, Application US/09103840A  
/ Patent No. 6284338  
/ GENERAL INFORMATION:  
/ APPLICANT: FLEISCHMAN, Robert D.  
/ APPLICANT: WHITE, Owen R.  
/ APPLICANT: FRASER, Claire M.  
/ APPLICANT: VENTER, John C.  
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
/ FILE REFERENCE: 24366-20007.00  
/ CURRENT APPLICATION NUMBER: US/09/103,840A  
/ CURRENT FILING DATE: 1998-06-24  
/ NUMBER OF SEQ ID NOS: 2  
/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 2  
/ LENGTH: 4403765  
/ TYPE: DNA  
/ ORGANISM: Mycobacterium tuberculosis  
/ FEATURE:  
/ OTHER INFORMATION: CDC 1551  
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2

Query Match 58.8%; Score 19.4; DB 3; Length 4403765;  
Best Local Similarity 74.2%; Pred. No. 72;  
Matches 23; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 3 CTGCAAACTCGTGTCTCCAGCATGCAAGG 33  
Db 2096169 CGGCAACCGGAGGTCTCTCCGAGTCCGGG 2096139

RESULT 10  
US-09-103-840A-1/c  
/ Sequence 1, Application US/09103840A  
/ Patent No. 6284338  
/ GENERAL INFORMATION:  
/ APPLICANT: FLEISCHMAN, Robert D.  
/ APPLICANT: WHITE, Owen R.  
/ APPLICANT: FRASER, Claire M.  
/ APPLICANT: VENTER, John C.  
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
/ FILE REFERENCE: 24366-20007.00  
/ CURRENT APPLICATION NUMBER: US/09/103,840A  
/ CURRENT FILING DATE: 1998-06-24  
/ NUMBER OF SEQ ID NOS: 2  
/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 1  
/ LENGTH: 4411529  
/ TYPE: DNA  
/ ORGANISM: Mycobacterium tuberculosis  
/ OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 58.8%; Score 19.4; DB 3; Length 4411529;  
Best Local Similarity 74.2%; Pred. No. 72;

Matches 23; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
Qy 3 CTGCAAACTCGTGTCTCCAGCATGCAAGG 33  
Db 2098979 CGGCAACCGGAGGTCTCTCCGAGTCCGGG 2098949

RESULT 11  
US-09-641-638-358  
/ Sequence 358, Application US/09641638  
/ Patent No. 6432648  
/ GENERAL INFORMATION:  
/ APPLICANT: Blumenfeld, Marta  
/ APPLICANT: Bougueleret, Lydie  
/ APPLICANT: Chumakov, Ilya  
/ APPLICANT: Cohen, Annick  
/ TITLE OF INVENTION: BIALLIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
/ FILE REFERENCE: GENSET.051CP1  
/ CURRENT APPLICATION NUMBER: US/09/641,638  
/ CURRENT FILING DATE: 2000-08-16  
/ PRIOR APPLICATION NUMBER: US 09/502,330  
/ PRIOR FILING DATE: 2000-02-11  
/ PRIOR APPLICATION NUMBER: US 60/133,200  
/ PRIOR FILING DATE: 1999-05-07  
/ PRIOR APPLICATION NUMBER: US 09/275,267  
/ PRIOR FILING DATE: 1999-03-23  
/ PRIOR APPLICATION NUMBER: US 60/119,917  
/ PRIOR FILING DATE: 1999-02-12  
/ NUMBER OF SEQ ID NOS: 1304  
/ SOFTWARE: Patent.in  
/ SEQ ID NO 358  
/ LENGTH: 1001  
/ TYPE: DNA  
/ ORGANISM: Homo Sapiens  
/ FEATURE:  
/ NAME/KEY: allele  
/ LOCATION: 501  
/ OTHER INFORMATION: 12-866-423 : polymorphic base C or T  
/ NAME/KEY: misc\_binding  
/ LOCATION: 482..500  
/ OTHER INFORMATION: 12-866-423.mis1  
/ NAME/KEY: misc\_binding  
/ LOCATION: 502..521  
/ OTHER INFORMATION: 12-866-423.mis2, potential complement  
/ NAME/KEY: primer\_bind  
/ LOCATION: 79..98  
/ OTHER INFORMATION: upstream amplification primer  
/ NAME/KEY: primer\_bind  
/ LOCATION: 589..609  
/ OTHER INFORMATION: downstream amplification primer, complement  
/ NAME/KEY: misc\_binding  
/ LOCATION: 489..513  
/ OTHER INFORMATION: 12-866-423 potential probe  
US-09-641-638-358

Query Match 58.2%; Score 19.2; DB 4; Length 1001;  
Best Local Similarity 80.8%; Pred. No. 36;  
Matches 21; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CTGCAAACTCGTGTCTCCAGCATG 28  
Db 906 CTGCAAACTCGTGTCTCCAGCATG 931

RESULT 12  
US-09-641-638-359  
/ Sequence 359, Application US/09641638  
/ Patent No. 6432648  
/ GENERAL INFORMATION:  
/ APPLICANT: Blumenfeld, Marta  
/ APPLICANT: Bougueleret, Lydie  
/ APPLICANT: Chumakov, Ilya

```

1  APPLICANT: Cohen, Anick
2  TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
3  TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
4  FILE REFERENCE: GENSET.05ICPI
5  CURRENT APPLICATION NUMBER: US/09/641,638
6  PRIOR FILING DATE: 2000-08-16
7  PRIOR APPLICATION NUMBER: US 09/502,330
8  PRIOR FILING DATE: 2000-02-11
9  PRIOR APPLICATION NUMBER: US 60/133,200
10 PRIOR FILING DATE: 1999-05-07
11 PRIOR APPLICATION NUMBER: US 09/275,267
12 PRIOR FILING DATE: 1999-03-23
13 PRIOR APPLICATION NUMBER: US 60/119,917
14 PRIOR FILING DATE: 1999-02-12
15 NUMBER OF SEQ ID NOS: 1304
16 SOFTWARE: Patent.pm
17 SEQ ID NO 359
18 LENGTH: 1001
19 TYPE: DNA
20 ORGANISM: Homo Sapiens
21 FEATURE:
22 NAME/KEY: allele
23 LOCATION: 501
24 OTHER INFORMATION: 12-867-47 : polymorphic base C or T
25 NAME/KEY: misc_binding
26 LOCATION: 481..500
27 OTHER INFORMATION: 12-867-47.misl, potential
28 NAME/KEY: misc_binding
29 LOCATION: 502..521
30 OTHER INFORMATION: 12-867-47.mis2, potential complement
31 NAME/KEY: primer_bind
32 LOCATION: 455..474
33 OTHER INFORMATION: upstream amplification primer
34 NAME/KEY: primer_bind
35 LOCATION: 898..918
36 OTHER INFORMATION: downstream amplification primer, complement
37 NAME/KEY: misc_binding
38 LOCATION: 489..513
39 OTHER INFORMATION: 12-867-47 potential probe
40 US-09-641-638-359

```

	Query Match	Similarity	Score	DB	Length
	Best Local	80.8%	Pred. No. 36		
Matches	21	Conservative	1	Mismatches	4
				Indels	0
				Gaps	0
Oy	3	CTGCAACTCGTGCTCTCCAGCATG	28		
Db	466	CTGCAAACTGTGTTCTCCAGCATG	491		

```

RESULT 13
US-09-904-615-27/c
: Sequence 27. Application US/09904615
: Patent No. 6566325
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: 49 Human Secreted Proteins
: FILE REFERENCE: P2032P1
: CURRENT APPLICATION NUMBER: US/09/904,615
: CURRENT FILING DATE: 2001-07-16
: PRIOR APPLICATION NUMBER: 09/511,554
: PRIOR FILING DATE: 2000-02-23
: PRIOR APPLICATION NUMBER: 60/097,917
: PRIOR FILING DATE: 1998-08-25
: PRIOR APPLICATION NUMBER: 60/098,634
: PRIOR FILING DATE: 1998-08-31
: NUMBER OF SEQ ID NOS: 170
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 27
: LENGTH: 1656
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-904-615-27

```

Query March	57.0%;	Score 18.8;	DB 4;	Length 1656;
Best Local Similarity	71.9%;	Pred. No. 58;		
Matches 23;	Conservative 1;	Mismatches 8;	Indels 0;	Gaps 0;

```

QY      2  CCTGCAACTCGTGTCTCTCCAGCATGCAGGG  33
          |||||
DB      231  CGTCAAACTCAGCGGTTCCAGCATGCAGTG  200

```

```

, RESULT 14
, US-09-214-278-4/c
, Sequence 4, Application US/09214278
, Patent No. 6291210
, GENERAL INFORMATION:
, APPLICANT: Sakano, Sei'ji
, APPLICANT: Itoh, Akira
, TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
, FILE REFERENCE: KP-8576
, CURRENT APPLICATION NUMBER: US/09/214,278
, CURRENT FILING DATE: 1999-01-26
, NUMBER OF SEQ ID NOS: 32
, SOFTWARE: PatentIn Ver. 2.1
, SEQ ID NO 4
, LENGTH: 3955

```

```

; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: CDS
; LOCATION: (12)..(3725)
; NAME/KEY: sig_peptide
; LOCATION: (12)..(89)
; NAME/KEY: mat_peptide
; LOCATION: (90)..(3725)
;
US-09-214-278-4

```

Query Match	57.0%;	Score 18.8;	DB 3;	length 3955;
Best Local Similarity	71.9%;	Pred. No. 67;		
Matches 23;	Conservative 1;	Mismatches 8;	Indels 0;	Gaps 0;

Qy 1 CCCTGCAAACTCGTGTCTCTCCAGCATGCAGG 32  
 Db 1298 CCCTTCAACTCATTTGGCTCCAGCTGCGCAGG 1267

RESULT 15  
 : Sequence 7, Application US/08400159  
 : Patent No. 5869282  
 : GENERAL INFORMATION:  
 : APPLICANT: Ish-Horowitz, David  
 : APPLICANT: Henriquez, Domingos M.P.  
 : APPLICANT: Lewis, Julian H.  
 : APPLICANT: Myat, Anna M.  
 : APPLICANT: Fleming, Robert J.  
 : APPLICANT: Artavanis-Tsakonas, Spyridon  
 : APPLICANT: Mann, Robert S.  
 : APPLICANT: Gray, Grace E.  
 : TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
 : TYPE OF INVENTION: SRRATE GENE AND METHODS BASED THEREON  
 : NUMBER OF SEQUENCES: 20  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Pennie & Edmonds  
 : STREET: 1155 Avenue of the Americas  
 : CITY: New York  
 : STATE: New York  
 : COUNTRY: USA  
 : ZIP: 10036-2711  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.30



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:14:25 ; Search time 119.257 Seconds  
(without alignments)  
680.690 Million cell updates/sec

Title: US-09-594-065-4

Perfect score: 33  
Sequence: 1 cccctgcaactcgtgkctccagcagcaggg 33

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1660708 seqs, 1229959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/CTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US00\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US00\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22.6	68.5	154746	12	US-09-827-688-8 Sequence 8, Appli
2	22.4	67.9	639	13	US-10-027-632-39522 Sequence 39522, A
3	22.4	67.9	639	13	US-10-027-632-39523 Sequence 39523, A
4	22.4	67.9	2815	10	US-09-312-762A-7 Sequence 7, Appli
5	20.2	61.2	3059	12	US-10-007-926A-408 Sequence 408, App
6	20	60.6	916	14	US-10-198-846-3325 Sequence 4325, App
7	19.8	60.0	3207	12	US-10-274-583-13 Sequence 13, Appli
8	19.8	60.0	3227	12	US-10-274-583-9 Sequence 9, Appli
9	19.8	60.0	3227	12	US-10-274-583-10 Sequence 10, Appli
10	19.8	60.0	3679	10	US-09-909-320-244 Sequence 244, App
11	19.8	60.0	3679	10	US-09-909-088B-244 Sequence 244, App
12	19.8	60.0	3679	10	US-09-905-291A-244 Sequence 244, App
13	19.8	60.0	3679	10	US-09-902-853-244 Sequence 244, App
14	19.8	60.0	3679	10	US-09-907-824-244 Sequence 244, App
15	19.8	60.0	3679	10	US-09-907-841-244 Sequence 244, App
16	19.8	60.0	3679	11	US-09-904-011-244 Sequence 244, App

C 17	19.8	60.0	3679	11	US-09-906-742-244	Sequence 244, App
C 18	19.8	60.0	3679	11	US-09-906-838-244	Sequence 244, App
C 19	19.8	60.0	3679	11	US-09-907-613-244	Sequence 244, App
C 20	19.8	60.0	3679	11	US-09-907-942-244	Sequence 244, App
C 21	19.8	60.0	3679	11	US-09-904-859-244	Sequence 244, App
C 22	19.8	60.0	3679	11	US-09-909-204-244	Sequence 244, App
C 23	19.8	60.0	3679	11	US-09-904-820-244	Sequence 244, App
C 24	19.8	60.0	3679	11	US-09-904-786-244	Sequence 244, App
C 25	19.8	60.0	3679	11	US-09-906-646-244	Sequence 244, App
C 26	19.8	60.0	3679	11	US-09-906-700-244	Sequence 244, App
C 27	19.8	60.0	3679	11	US-09-903-786-244	Sequence 244, App
C 28	19.8	60.0	3679	11	US-09-902-903-244	Sequence 244, App
C 29	19.8	60.0	3679	11	US-09-903-749A-244	Sequence 244, App
C 30	19.8	60.0	3679	11	US-09-904-119-244	Sequence 244, App
C 31	19.8	60.0	3679	11	US-09-904-956-244	Sequence 244, App
C 32	19.8	60.0	3679	11	US-09-902-736-244	Sequence 244, App
C 33	19.8	60.0	3679	11	US-09-907-794-244	Sequence 244, App
C 34	19.8	60.0	3679	11	US-09-903-943-244	Sequence 244, App
C 35	19.8	60.0	3679	11	US-09-904-562-244	Sequence 244, App
C 36	19.8	60.0	3679	11	US-09-907-925-244	Sequence 244, App
C 37	19.8	60.0	3679	11	US-09-902-692-244	Sequence 244, App
C 38	19.8	60.0	3679	11	US-09-903-520-244	Sequence 244, App
C 39	19.8	60.0	3679	11	US-09-905-056-244	Sequence 244, App
C 40	19.8	60.0	3679	11	US-09-909-064-244	Sequence 244, App
C 41	19.8	60.0	3679	11	US-09-904-553-244	Sequence 244, App
C 42	19.8	60.0	3679	11	US-09-905-181-244	Sequence 244, App
C 43	19.8	60.0	3679	11	US-09-905-088-244	Sequence 244, App
C 44	19.8	60.0	3679	11	US-09-907-575-244	Sequence 244, App
C 45	19.8	60.0	3679	11	US-09-905-075-244	Sequence 244, App

## ALIGNMENTS

RESULT 1  
US-09-827-688-8  
; Sequence 8, Application US/09827688  
; Publication No. US20030165476A1  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BERRA  
; APPLICANT: BHOQAL, BALBIR  
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION 1  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: P01949US/10004014  
; CURRENT APPLICATION NUMBER: US/09/827,688  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8

Query Match 68.5%; Score 22.6; DB 12; Length 154746;  
Best Local Similarity 95.7%; Pred. No. 4.3;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 CAAACTGCTGCTCCAGCATG 28  
DB 54077 CAAACTGCTGCTCCAGCATG 54099

RESULT 2  
US-10-027-632-39522  
; Sequence 39522, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39522
LENGTH: 639
TYPE: DNA
ORGANISM: Human
US-10-027-632-39522
```

Query Match 67.9%; Score 22.4; DB 13; Length 639;  
Best Local Similarity 88.5%; Pred. No. 4.7;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
4 TCGAACTGCTGCTCTCCAGCATGC 29
|||||
581 TCGAACTGCTGCTCTCCAGCATGC 606
```

```
RESULT 3
US-10-027-632-39523
Sequence 39523, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39523
LENGTH: 639
TYPE: DNA
ORGANISM: Human
US-10-027-632-39523
```

Query Match 67.9%; Score 22.4; DB 13; Length 639;  
Best Local Similarity 88.5%; Pred. No. 4.7;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
4 TCGAACTGCTGCTCTCCAGCATGC 29
|||||
581 TCGAACTGCTGCTCTCCAGCATGC 606
```

```
RESULT 4
US-09-312-762A-7/c
Sequence 7, Application US/09312762A
Patent No. US20020115069A1
GENERAL INFORMATION:
APPLICANT: MIA HOROWITZ ET AL.
TITLE OF INVENTION: EH DOMAIN CONTAINING GENES AND PROTEINS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
STREET: 2001 Jefferson Davis Highway, Suite 207
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
COMPUTER: Twinhead* Slimnote-890TX
OPERATING SYSTEM: MS DOS version 6.2,
OPERATING SYSTEM: Windows version 3.11
SOFTWARE: Word for Windows version 2.0 converted to
SOFTWARE: an ASCII file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/312,762A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/026,898
FILING DATE: 20 FEB 1998
ATTORNEY/AGENT INFORMATION:
NAME: Friedman, Mark M.
REGISTRATION NUMBER: 33,883
REFERENCE/DOCKET NUMBER: 916/10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 972-3-562553
TELEFAX: 972-3-562554
TELEX:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2815
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-312-762A-7
```

Query Match 67.9%; Score 22.4; DB 10; Length 2815;  
Best Local Similarity 88.5%; Pred. No. 4.8;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
4 TCGAACTGCTGCTCTCCAGCATGC 29
|||||
580 TCGAACTGCTGCTCTCCAGCATGC 955
```

```
RESULT 5
US-10-007-926A-408
Sequence 408, Application US/10007926A
Publication No. US2003014359A1
GENERAL INFORMATION:
APPLICANT: BERTUCCI, FRANCOIS
APPLICANT: HOUJAGATTE, REMI
APPLICANT: BIRNBAUM, DANIEL
APPLICANT: NGUYEN, CATHERINE
APPLICANT: VIENS, PATRICE
APPLICANT: FERT, VINCENT
TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
FILE REFERENCE: 1546-R-00
CURRENT APPLICATION NUMBER: US/10/007,926A
CURRENT FILING DATE: 2001-12-07
PRIOR APPLICATION NUMBER: 60/254,090
```



PRIOR FILING DATE: 2000-12-08  
NUMBER OF SEQ ID NOS: 468  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 408  
LENGTH: 3059  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Interleukin enhancer binding factor 1 (ILF1)  
US-10-007-926A-408

Query Match 61.2% Score 20.2; DB 12; Length 3059;  
Best Local Similarity 75.8% Pred. No. 42;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 CCTGCAAACTCGTCTCTCCAGTCAGG 33  
Db 797 CGCTGCAAACTCTGCTCCAGCCCGGAGG 829

RESULT 6  
US-10-198-846-4325

Sequence 4325, Application US/10198846  
Publication No. US2003009974A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinhmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4325  
LENGTH: 916  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 439, 451, 525, 571, 596, 627, 636, 642, 668, 686, 689, 706,  
LOCATION: 711, 715, 740, 759, 779, 782, 783, 789, 790, 796, 798, 802,  
LOCATION: 812, 816, 817, 826, 828, 831, 835, 836, 837, 841, 842, 846,  
LOCATION: 850, 856, 871, 872, 873, 876, 879, 880, 882, 883, 889  
OTHER INFORMATION: n = A,T,C or G  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 890, 895, 896, 897, 898, 900, 901, 903, 904, 909, 910, 911,  
LOCATION: 914  
OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-4325

Query Match 60.6% Score 20; DB 14; Length 916;  
Best Local Similarity 76.7% Pred. No. 49;  
Matches 23; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 4 TGCAACCTGCTCTCCAGTCAGG 33  
Db 486 TGCAACCTGCTCTCCAGTCAGG 515

RESULT 7  
US-10-274-583-13/c  
Sequence 13, Application US/10274583  
Publication No. US20030138431A1  
GENERAL INFORMATION:  
APPLICANT: Exelixis, Inc.

TITLE OF INVENTION: LRCAPS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE  
FILE REFERENCE: EX02-119C  
CURRENT APPLICATION NUMBER: US/10/274,583  
CURRENT FILING DATE: 2002-10-21  
PRIOR APPLICATION NUMBER: 60/338,733  
PRIOR FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: 60/357,600  
PRIOR FILING DATE: 2002-02-15  
PRIOR APPLICATION NUMBER: 60/361,196  
PRIOR FILING DATE: 2002-03-01  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 13  
LENGTH: 3207  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-274-583-13

Query Match 60.0% Score 19.8; DB 12; Length 3207;  
Best Local Similarity 84.0% Pred. No. 61;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CCTGCAAACTCGTCTCTCCAGC 25  
Db 793 CCTGCAAACTGCTCTCTCCAGC 769

RESULT 8  
US-10-274-583-9/c

Sequence 9, Application US/10274583  
Publication No. US20030138431A1  
GENERAL INFORMATION:  
APPLICANT: Exelixis, Inc.  
TITLE OF INVENTION: LRCAPS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE  
FILE REFERENCE: EX02-119C  
CURRENT APPLICATION NUMBER: US/10/274,583  
CURRENT FILING DATE: 2002-10-21  
PRIOR APPLICATION NUMBER: 60/338,733  
PRIOR FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: 60/357,600  
PRIOR FILING DATE: 2002-02-15  
PRIOR APPLICATION NUMBER: 60/361,196  
PRIOR FILING DATE: 2002-03-01  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 9  
LENGTH: 3227  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-274-583-9

Query Match 60.0% Score 19.8; DB 12; Length 3227;  
Best Local Similarity 84.0% Pred. No. 61;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CCTGCAAACTCGTCTCTCCAGC 25  
Db 805 CCTGCAAACTGCTCTCTCCAGC 781

RESULT 9  
US-10-274-583-10/c  
Sequence 10, Application US/10274583  
Publication No. US20030138431A1  
GENERAL INFORMATION:  
APPLICANT: Exelixis, Inc.  
TITLE OF INVENTION: LRCAPS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE  
FILE REFERENCE: EX02-119C  
CURRENT APPLICATION NUMBER: US/10/274,583  
CURRENT FILING DATE: 2002-10-21  
PRIOR APPLICATION NUMBER: 60/338,733  
PRIOR FILING DATE: 2001-10-22  
PRIOR APPLICATION NUMBER: 60/357,600

```
/ PRIOR FILING DATE: 2002-02-15
/ PRIOR APPLICATION NUMBER: 60/161,196
/ PRIOR FILING DATE: 2002-03-01
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 10
/ LENGTH: 3227
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-274-583-10

Query Match      60.0%; Score 19.8; DB 12; Length 3227;
Best Local Similarity 84.0%; Pred. No. 61;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 CCTGCAAACTGCTGCTCTCCAGC 25
DB      805 CCTGCAAGCTGTGTCTCTCCAGC 781

RESULT 10
US-09-909-320-244/C
/ Sequence 244, Application US/09909320
/ Patent No. US20020132240A1
/ GENERAL INFORMATION:
/ APPLICANT: Genentech, Inc.
/ APPLICANT: Ashkenazi, Avi
/ APPLICANT: Botstein, David
/ APPLICANT: Desnovers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, A.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Hillan, Kenneth, J.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Mather, Jennie P.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ TITLE OF INVENTION: Acids Encoding the Same
/ FILE REFERENCE: 10466-14
/ CURRENT APPLICATION NUMBER: US/09/909,320
/ CURRENT FILING DATE: 2002-01-04
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
```

```
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 244
/ LENGTH: 3679
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-09-909-320-244

Query Match      60.0%; Score 19.8; DB 10; Length 3679;
Best Local Similarity 84.0%; Pred. No. 62;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 CCTGCAAACTGCTGCTCTCCAGC 25
DB      1297 CCTGCAAGCTGTGTCTCTCCAGC 1273

RESULT 11
US-09-909-0888-244/C
/ Sequence 244, Application US/099090888
/ Patent No. US20020146709A1
/ GENERAL INFORMATION:
/ APPLICANT: Genentech, Inc.
/ APPLICANT: Ashkenazi, Avi
/ APPLICANT: Botstein, David
/ APPLICANT: Desnovers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Gao, Wei-Qiang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerritsen, Mary E.
/ APPLICANT: Goddard, A.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Hillan, Kenneth, J.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Mather, Jennie P.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ TITLE OF INVENTION: Acids Encoding the Same
/ FILE REFERENCE: 10466-14
/ CURRENT APPLICATION NUMBER: US/09/909,0888
/ CURRENT FILING DATE: 2001-07-18
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
```

```
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 244
;; LENGTH: 3679
;; TYPE: DNA
;; ORGANISM: Homo Sapien
US-09-909-088B-244
```

```
Query Match      60.0%; Score 19.8; DB 10; Length 3679;
Best Local Similarity 84.0%; Pred. No. 62;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1  CCCTGCAGAACTCGTGTCTCTCAGC 25
Db      1297 CCCTGCAGAACTCGTGTCTCTCAGC 1273
```

```
RESULT 12
US-09-905-291A-244/C
; Sequence 244, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICATION: Genentech, Inc.
; APPLICATION: Ashkenazi, Avi
; APPLICATION: Botstein, David
; APPLICATION: Deenoyers, Luc
; APPLICATION: Baton, Dan L.
; APPLICATION: Ferrara, Napoleone
; APPLICATION: Filvaroff, Ellen
; APPLICATION: Fong, Sherman
; APPLICATION: Gao, Wei-Qiang
; APPLICATION: Gerber, Hanspeter
; APPLICATION: Gerltsen, Mary E.
; APPLICATION: Goddard, A.
; APPLICATION: Grimaldi, Paul J.
; APPLICATION: Gurney, Austin L.
; APPLICATION: Hillan, Kenneth, J.
; APPLICATION: Kijavlin, Ivar J.
; APPLICATION: Macher, Jennie P.
; APPLICATION: Pan, James
; APPLICATION: Paoni, Nicholas F.
; APPLICATION: Roy, Margaret Ann
; APPLICATION: Stewart, Timothy A.
; APPLICATION: Tumas, Daniel
; APPLICATION: Williams, P. Mickey
```

```
;; APPLICATION: Wood, William, I.
;; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;; FILE REFERENCE: 10466-14
;; CURRENT APPLICATION NUMBER: US/09/905,291A
;; CURRENT FILING DATE: 2001-07-12
;; PRIOR APPLICATION NUMBER: PCT/US00/04414
;; PRIOR FILING DATE: 2000-02-22
;; PRIOR APPLICATION NUMBER: US 60/143,048
;; PRIOR FILING DATE: 1999-07-07
;; PRIOR APPLICATION NUMBER: US 60/145,698
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/146,222
;; PRIOR FILING DATE: 1999-07-28
;; PRIOR APPLICATION NUMBER: PCT/US99/20594
;; PRIOR FILING DATE: 1999-09-08
;; PRIOR APPLICATION NUMBER: PCT/US99/20944
;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: PCT/US99/21090
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/21547
;; PRIOR FILING DATE: 1999-09-15
;; PRIOR APPLICATION NUMBER: PCT/US99/23089
;; PRIOR FILING DATE: 1999-10-05
;; PRIOR APPLICATION NUMBER: PCT/US99/28214
;; PRIOR FILING DATE: 1999-11-29
;; PRIOR APPLICATION NUMBER: PCT/US99/28313
;; PRIOR FILING DATE: 1999-11-30
;; PRIOR APPLICATION NUMBER: PCT/US99/28564
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/28565
;; PRIOR FILING DATE: 1999-12-02
;; PRIOR APPLICATION NUMBER: PCT/US99/30095
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: PCT/US99/30911
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US99/30999
;; PRIOR FILING DATE: 1999-12-20
;; PRIOR APPLICATION NUMBER: PCT/US00/00219
;; PRIOR FILING DATE: 2000-01-05
;; NUMBER OF SEQ ID NOS: 423
;; SEQ ID NO 244
;; LENGTH: 3679
;; TYPE: DNA
;; ORGANISM: Homo Sapien
US-09-905-291A-244
```

```
Query Match      60.0%; Score 19.8; DB 10; Length 3679;
Best Local Similarity 84.0%; Pred. No. 62;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      1  CCCTGCAGAACTCGTGTCTCTCAGC 25
Db      1297 CCCTGCAGAACTCGTGTCTCTCAGC 1273
```

```
RESULT 13
US-09-902-853-244/C
; Sequence 244, Application US/09902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICATION: Genentech, Inc.
; APPLICATION: Ashkenazi, Avi
; APPLICATION: Botstein, David
; APPLICATION: Deenoyers, Luc
; APPLICATION: Baton, Dan L.
; APPLICATION: Ferrara, Napoleone
; APPLICATION: Filvaroff, Ellen
; APPLICATION: Fong, Sherman
; APPLICATION: Gao, Wei-Qiang
; APPLICATION: Gerber, Hanspeter
; APPLICATION: Gerltsen, Mary E.
; APPLICATION: Goddard, A.
```

```

/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Hillan, Kenneth J.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Mather, Jennie P.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: 10466-14
/ CURRENT APPLICATION NUMBER: US/09/902,853
/ CURRENT FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: US/09/665,350
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-09
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 244
/ LENGTH: 3679
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-09-902-853-244

Query Match      60.0%; Score 19.8; DB 10; Length 3679;
Best Local Similarity 84.0%; Pred. No. 62;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 CCCTGCAAACTGCTGCTCTCCAGC 25
DB      1297 CCCTGCAAGCTGCTGCTCCAGC 1273

RESULT 14
US-09-907-824-244/c
/ Sequence 244, Application US/09907824
/ Publication No. US20020197671A1
/ GENERAL INFORMATION:
```

```

/ APPLICANT: Genentech, Inc.
/ APPLICANT: Ashkenazi, Avi
/ APPLICANT: Botsstein, David
/ APPLICANT: Deemoyers, Luc
/ APPLICANT: Eaton, Dan L.
/ APPLICANT: Ferrara, Napoleone
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Fong, Sherman
/ APPLICANT: Geo, Wei-Olang
/ APPLICANT: Gerber, Hanspeter
/ APPLICANT: Gerritsen, Mary B.
/ APPLICANT: Goddard, A.
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Grimaldi, Christopher J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Hillan, Kenneth, J.
/ APPLICANT: Kijavlin, Ivar J.
/ APPLICANT: Mather, Jennie P.
/ APPLICANT: Pan, James
/ APPLICANT: Paoni, Nicholas F.
/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ FILE REFERENCE: 10466-14
/ CURRENT APPLICATION NUMBER: US/09/907,824
/ CURRENT FILING DATE: 2001-07-17
/ PRIOR APPLICATION NUMBER: 09/665,350
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: PCT/US00/04414
/ PRIOR FILING DATE: 2000-02-22
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 244
/ LENGTH: 3679
/ TYPE: DNA
/ ORGANISM: Homo Sapien
US-09-907-824-244
```

Query Match 60.0%; Score 19.8; DB 10; Length 3679;  
Best Local Similarity 84.0%; Pred. No. 62;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTGCAACTCGTGTCTCCAGC 25  
DB 1297 CCTGCAAGCTGTGTCTCCAGC 1273

## RESULT 15

US-09-907-841-244/C  
Sequence 244, Application US/09907841  
Publication No. US20020198366A1

## GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
APPLICANT: Ashtenazi, Avi  
APPLICANT: Borstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gerritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kijavlin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
TITLE OF INVENTION: Acids Encoding the Same  
FILE REFERENCE: 1046-14  
CURRENT APPLICATION NUMBER: US/09/907,841  
CURRENT FILING DATE: 2001-11-20  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 244  
LENGTH: 3679

TYPE: DNA

ORGANISM: Homo Sapien

US-09-907-841-244

Query Match 60.0%; Score 19.8; DB 10; Length 3679;

Best Local Similarity 84.0%; Pred. No. 62;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCTGCAACTCGTGTCTCCAGC 25  
DB 1297 CCTGCAAGCTGTGTCTCCAGC 1273

Search completed: September 16, 2003, 22:54:39  
Job time : 120.257 secs

**BEST AVAILABLE COPY**







/note="unknown EST (GB|BEJ34066, evidence: BLASTN, 97%, match=403)"

BASE COUNT 1159 a 794 c 778 g 1132 t

ORIGIN

Query Match 71.5%; Score 23.6; DB 11; Length 3863;  
 Best Local Similarity 81.2%; Pred. No. 2.8e+02;  
 Matches 26; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 CCTGCAACTCGTGTCTCCAGCATGCAG 32  
 |||||  
 1433 CCCAGCAACTCGTGTCTACACAAATGCAG 1464

RESULT 3  
 CBI59107/c 590 bp mRNA linear EST 30-JAN-2003  
 LOCUS K-EST0218577 L18POOL1n1 Homo sapiens CDNA clone L18POOL1n1-37-E05  
 DEFINITION 5', mRNA sequence.  
 ACCESSION CBI59107  
 VERSION CBI59107.1 GI:28145233  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 590)  
 AUTHORS Oh, K.Y., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
 Kim, Y.S.  
 TITLE 21c Frontier Korean EST Project 2001  
 JOURNAL Unpublished  
 COMMENT Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongsung@mail.kribb.re.kr  
 Plate: 37 row: B column: 05  
 High quality sequence stop: 590.  
 Location/Qualifiers  
 1..590  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="L18POOL1n1-37-E05"  
 /cell\_line="SNU-354+Cho-CK+Choi-CK+HLK-3"  
 /lab\_host="Top10P"  
 /clone\_lib="L18POOL1n1"  
 /note="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;  
 laboratory and it was constructed by the Soares  
 laboratory and it was constructed as described by Bonaldo,  
 M.F., Lennon, G. and Soares, M.B. (1996), Genome Research  
 6(9): 791-806. RNA was prepared from harvested cell  
 culture."

BASE COUNT 146 a 158 c 187 g 99 t

ORIGIN

Query Match 67.9%; Score 22.4; DB 14; Length 590;  
 Best Local Similarity 88.5%; Pred. No. 4.5e+02;  
 Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 TGCAAACTCGTGTCTCCAGCATGC 29  
 |||||  
 359 TGCAAACTCGTGTCTCCAGCATGC 334

RESULT 4  
 AA972967 605 bp mRNA linear EST 07-JUL-1998  
 LOCUS ops5d11.e1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens CDNA clone  
 IMAGE:1577877 3' similar to TR:014611 O14611 HPAST.; contains

element TARI repetitive element ;, mRNA sequence.

ACCESSION AA972967  
 VERSION AA972967.1 GI:3148147  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 605)  
 AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CCGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Insert Length: 932 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 444.  
 Location/Qualifiers  
 1..605  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1577877"  
 /lab\_host="VDH10B"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
 Equal amounts of plasmid DNA from three normalized  
 libraries (fetal lung NDHL19W, testis NHT, and B-cell  
 NCI CGAP GCB1) were mixed, and ss circles were made in  
 vitro. Following HAP purification, this DNA was used as  
 tracer in a subtractive hybridization reaction. The driver  
 was PCR-amplified cDNAs from pools of 5,000 clones made  
 from the same 3 libraries. The pools consisted of  
 I.M.A.G.E. clones 297480-302087, 682632-687239,  
 726408-728711, and 729096-731399. Subtraction by Bento  
 Soares and M. Patricia Bonaldo."

BASE COUNT 96 a 194 c 164 g 151 t

ORIGIN

Query Match 67.9%; Score 22.4; DB 9; Length 605;  
 Best Local Similarity 88.5%; Pred. No. 4.5e+02;  
 Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 TGCAAACTCGTGTCTCCAGCATGC 29  
 |||||  
 222 TGCAAACTCGTGTCTCCAGCATGC 247

RESULT 5  
 HSM096537/c standard; RNA; EST; 639 BP.  
 ID HSM096537  
 AC BX502820;  
 XX BX502820.1  
 SV  
 XX 09-MAY-2003 (Rel. 75, Created)  
 DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)  
 XX  
 DE Homo sapiens mRNA; EST DKFZp779E0170\_r1 (from clone DKFZp779E0170)  
 EST; expressed sequence tag.  
 XX  
 KW  
 XX Homo sapiens (human)  
 OS  
 XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 XX  
 XX [1]  
 RN 1-639  
 RP

RA	Bahr A., Lauber J., Mewes H.W., Weil B., Amid C., Osaenger A., Fobo G.,
RN	Han M., Wiemann S.;
RT	Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL	MISr, Ingelstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX	
CC	This is the 5' sequence of the clone insert
CC	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC	Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC	consortium by Qiagen (Hilden/Germany) within the cDNA sequencing
CC	consortium of the German Genome Project.
CC	No nt sequence available.
CC	This clone (DKFZp79B0170) is available at the RZPD in Berlin.
CC	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC	14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX	
FH	Key Location/Qualifiers
FT	
FT	source
FT	1. 639
FT	/db_xref=taxon:9606"
FT	/mol_type="mRNA"
FT	/organism="Homo sapiens"
FT	/clone="DKFZp79B0170"
FT	/clone_lib="779 (synonym: hmccl). Vector pSport1_Sfi; host:
FT	DH10B; sites SfiI + SfiIB"
FT	/dev_stage="fetal"
FT	/tissue_type="liver"
XX	
SQ	Sequence 639 BP; 160 A; 171 C; 194 G; 114 T; 0 other;
	Query Match 67.9%; Score 22.4; DB 2; Length 639;
	Best Local Similarity 88.5%; Pred. No. 4,6e+02;
	Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY	4 TGCAACTCGAGTCTCCTCCAGCATGC 29
	:
Db	278 TGCAACTCGTCTGTCGTCCAGATGC 253
RESULT 6	
CBI55818/c	639 bp mRNA linear EST 29-JAN-2003
LOCUS	K-EST021433 B2N807043 Homo sapiens CDNA clone B2N807043-16-C09 5'
DEFINITION	mRNA sequence.
ACCESSION	CBI55818
VERSION	CBI55818.1 GI:28140931
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 639)
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
	Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
	Kim,Y.S.
TITLE	21C Frontier Korean EST Project 2001
JOURNAL	unpublished
COMMENT	Contact: Kim YS
	Genome Research Center
	Korea Research Institute of Bioscience & Biotechnology
	52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
	Tel.: +82-42-860-4470
	Fax: +82-42-860-4409
	Email: yongsung@mail.kr.ibm.re.kr
	Plate: 16 row: C column: 09
	High quality sequence frop: 639.
FEATURES	Location/Qualifiers
source	1..639
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref=taxon:9606"
	/clone="B2N807043-16-C09"
	/sex="M"

```

/lab host="TopIOF"
/cldn=organs: Brain; Vector: PCNS-D2; Site: 1; Ecoref:
Site_2; Nct1: The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
inact cDNA was ligated with DNA-RNA linker including
Ecoref site by treatment of T4 RNA ligase and the first
strand cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E.coli DNA ligase after digestion of
Ecoref which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli TopIOF' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT      157 a      170 c      202 g      110 t
ORIGIN
Query Match     67.9%; Score 22.4; DB 14; Length 639;
Best Local Similarity 88.5%; Pred. No. 4.6e+02;
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      4 TGCAAACTCGTGTCCTCCAGCATGC 29
|||||
|||||
|||||
|||||
|||||
Db      350 TGCAAACTCGTGTCCTCCAGCATGC 325

RESULT 7
CA488926      884 bp      mRNA      linear      EST 14-NOV-2002
LOCUS CA488926/c
DEFINITION AGNCOCURT 10808120 MAPCL Homo sapiens CDNA clone IMAGE:6721132 5'.
ACCESSION CA488926
VERSION CA488926
KEYWORDS mRNA sequence.
SOURCE CA488926.1 GI:24951717
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 884)
NTH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Straubeberg, Ph.D.
Email: cgadps-iemail.nih.gov
Tissue Procurement: Kristi A. Egland, Ira Pastan
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC Clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LAM14281 row: m column: 04
High quality sequence stop: 664.
Location/Qualifiers
1..884
/organism="Homo sapiens"
/mol type="mRNA"
/db xref="taxon:9606"
/cldn="IMAGE:6721132"
/cell_line="ZR-75-1, MCP7, SK-BR-3, MDA-MB-231, hTERT-HME1
'LNCaP'
/lab host="EMDH10B"
/cldn=lib="MAPCL"
/note="Vector: PGMV-SPOrt6; Site_1: Ecoref; Site_2: Not 1;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dt. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Egland, James J. Vincent, Robert Straubeberg,
Bungkok Lee & Ira Pastan: Discovery of new breast

```

cancer genes encoding membrane and secreted proteins.

## BASE COUNT

238 a 231 c 216 g 199 t

## ORIGIN

Query Match 67.9%; Score 22.4; DB 14; Length 884;  
Best Local Similarity 88.5%; Pred. No. 5.1e+02;  
Matches 23; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

## Qy

4 TCGAAGCTGCTGCTCCAGCATGC 29

## Db

54 TCGAAGCTGCTGCTCCAGCATGC 29

RESULT 8  
LOCUS

BZ375631 759 bp DNA linear GSS 26-NOV-2002

1659907.5 genomic survey sequence.

## ACCESSION

BZ375631

## VERSION

BZ375631.1 GI:25463412

## KEYWORDS

GSS.

## SOURCE

Zea mays

## ORGANISM

Zea mays

## REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

## AUTHORS

Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dechhia, N., Karszenowicz, P., King, L., Miller, B., Muller, S., Naschimoto, L., Zuber, T., McCombie, W.R. and Martienssen, R.A.

## TITLE

Genomic shotgun sequences from Zea mays (methyl-filtered)

## JOURNAL

Unpublished

## COMMENT

Contact: W. Richard McCombie

## FEATURES

Location/Qualifiers

## SOURCE

1..759

## ORGANISM

"/organism="Zea mays"

## MOIETY

"/moiety="genomic DNA"

## CULTIVAR

"/cultivar="B73"

## DB XREF

"/db\_xref="taxon:4577"

## CLONE

"/clone="1659907"

## LAB HOST

"/lab\_host="DH5a"

## CLONE LIB

"/clone\_lib="WGS-ZmaysF (DH5a methyl filtered)"

## NOTE

"/note="Organ: immature ears; Site\_1: Xba I; Site\_2: Xba I; The vector was digested with XbaI and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, b/g reads in pUC19). The same ligation was transformed into DH5a."

## BASE COUNT

199 a 174 c 175 g 211 t

## ORIGIN

Query Match 67.3%; Score 22.2; DB 29; Length 759;  
Best Local Similarity 82.8%; Pred. No. 5.7e+02;  
Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

## Qy

2 CCTGCAAGCTGCTGCTCCAGCATGCA 30

## Db

542 CCTGCAAGCTGCTGCTCCAGCATGCA 570

## RESULT 9

BY540314/ 401 bp mRNA linear EST 14-DEC-2002

## LOCUS

BY540314 RIKEN full-length enriched, B6-derived C01 +ve dendritic

## DEFINITION

cells Mus musculus cDNA clone F730006E21.3, mRNA sequence.

## ACCESSION

BY540314

## VERSION

BY540314.1 GI:26874693

## KEYWORDS

EST.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## AUTHORS

Okazaki, Y., Furum, M., Kasekawa, T., Adachi, J., Bono, H., Kondo, S., Nishida, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarini, R., Hill, D.P., Bul, C., Hume, D.A., Quackenbush, J., Schmitt, L.M., Kanap, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lehnard, B., Lyons, P.A., Maglott, D.R., Maltas, K., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, M.J., Pereira, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sander, A., Schneider, C., Semp, C.A., Setou, M., Shimada, K., Sultana, R., Takemura, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlschmidt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyszynski, A., Yamaoka, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Maki, K., Kawai, J., Aizawa, K., Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, R. and Hayashizaki, Y.

## JOURNAL

Analysis of the mouse transcriptome based on functional annotation

## MEDLINE

Nature 420, 563-573 (2002)

## PUBMED

12466851

## COMMENT

Contact: Yoshihide Hayashizaki

## LABORATORY

Laboratory for Genome Exploration Research Group, RIKEN Genomic

## SCIENCES CENTER (GSC), YOKOHAMA INSTITUTE

The Institute of Physical and Chemical Research (RIKEN)

## 1-7-22 SUEHIRO-CHO, TSURUMI-KU, YOKOHAMA, KANAGAWA 230-0045, JAPAN

## TEL: 81-45-503-9222

## FAX: 81-45-503-9216

## EMAIL: genome-res@sc.riken.go.jp

## URL: http://genome.gsc.riken.go.jp/

## ALZAWA, K., AKIMURA, T., ARAKAWA, T.,

## T., IMOTANI, K., ISHII, Y., ITOH, M., KAWAI, J., KONNO, H., MIYAZAKI, A.,

## MURATA, M., NAKAMURA, M., NOMURA, K., NUMAZAKI, R., OHNO, M., SAKAI, K.,

## SAKAZUME, N., SAKAKI, D., SATO, K., SHIBATA, K., SHIRAKI, T., TAGAMI, M.,

## WAKI, K., WATANAKI, A., YAMAMOTO, M. and HAYASHIZAKI, Y. Direct

## SUBMISSION

## COMPUTATIONAL ANALYSIS OF FULL-LENGTH MOUSE CDNAS COMPARED WITH

## HUMAN GENOME SEQUENCES Mamm. Genome. 12, 673-677 (2001)

## NORMALIZATION AND SUBTRACTION OF CAP-TRAP-SELECTED CDNAS TO

## PREPARE FULL-LENGTH CDNA LIBRARIES FOR RAPID DISCOVERY OF NEW

## GENES. Genome Res. 10 (10), 1617-1630 (2000)

## RIKEN INTEGRATED SEQUENCE ANALYSIS (RISA) SYSTEM--384-FORMAT

## SEQUENCING PIPELINE WITH 384 MULTICAPILLARY SEQUENCER. Genome Res.

## 10 (11), 1757-1771 (2000)

## COMPUTER-BASED METHODS FOR THE MOUSE FULL-LENGTH CDNA

## ENCYCLOPEDIA: REAL-TIME SEQUENCE CLUSTERING FOR CONSTRUCTION OF A

## NONREdundant CDNA LIBRARY. Genome Res. 11 (2), 281-289 (2001)

## CDNA LIBRARY WAS PREPARED AND SEQUENCED IN MOUSE GENOME

## ENCYCLOPEDIA PROJECT OF GENOME EXPLORATION RESEARCH GROUP IN RIKEN

## GENOMIC SCIENCES CENTER AND GENOME SCIENCE LABORATORY IN RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

location/Qualifiers  
1..401  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="F73006E21"  
/cell\_type="B6-derived CD11 +ve dendritic cells"  
/clone\_1lb="RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells"  
BASE COUNT  
90 a 105 c 106 g 100 t  
ORIGIN

Query Match  
Best Local Similarity 78.1%; Score 22; DB 13; Length 401;  
Matches 25; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy  
1 CCTGCAAACTGCTGCTCTCCAGCATGCAGC 32  
37 CCTGCAATCTGCTGCTCTCCAGCATGCAGC 6

RESULT 10  
BY541787/c 405 bp mRNA linear EST 14-DEC-2002  
LOCUS  
DEFINITION  
BY541787 RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells Mus musculus cDNA clone F730022F15 3', mRNA sequence.  
BY541787  
1 GI:26876166

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 405)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bone, H., Kondo, S., Nikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamamaki, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schorbach, C., Gojodori, I., Balderelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matluda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusc, V., Chothia, C., Corbett, L. B., Cousins, S., Dalla, E., Dragan, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hinkawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedir, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Piliat, R., Pontius, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Sempel, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynchaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carlini, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Wacraton, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE  
JOURNAL  
MEDLINE  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
22354683

## PUBMED

12466851

Contact: Yoshitake Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Shukho-cho, Tsukuba, Ibaraki, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: [genome-gsc.riken.go.jp](mailto:genome-gsc.riken.go.jp)URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Aizawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multichannel sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics

Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome

Trust/MRC Building Addenbrookes Hospital Cambridge) whose

assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

## FEATURES

source

location/Qualifiers  
1..405  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="F730022F15"  
/cell\_type="B6-derived CD11 +ve dendritic cells"  
/clone\_1lb="RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells"  
BASE COUNT  
92 a 107 c 106 g 100 t  
ORIGIN

Query Match  
Best Local Similarity 78.1%; Score 22; DB 13; Length 405;  
Matches 25; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy  
1 CCTGCAAACTGCTGCTCTCCAGCATGCAGC 32  
37 CCTGCAATCTGCTGCTCTCCAGCATGCAGC 6

RESULT 11  
BY536425/c 439 bp mRNA linear EST 14-DEC-2002  
LOCUS  
DEFINITION  
BY536425 RIKEN full-length enriched, NOD-derived CD11c +ve dendritic cells Mus musculus cDNA clone F630310H17 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)

REFERENCE  
AUTHORS  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 439)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bone, H., Kondo, S., Nikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamamaki, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schorbach, C., Gojodori, I., Balderelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matluda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusc, V., Chothia, C., Corbett, L. B., Cousins, S., Dalla, E., Dragan, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Giasi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hinkawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedir, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Piliat, R., Pontius, J. U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Sempel, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynchaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carlini, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aizawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Wacraton, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

REFERENCE 1 (bases 1 to 439)  
 AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nishida,I., Osato,N., Saito,R., Suzuki,H., Yamana,H., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bull,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Brad,D., Brusch,V., Chotila,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimond,S., Guernicich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltas,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perle,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Saitana,R., Takenaka,Y., Taylor,M.S., Teasdale,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,M., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavalan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Kono,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shingawa,A., Yasuniishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.  
 ANALYSIS: The mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 NATURE 420, 563-573 (2002)  
 22354683  
 12466851  
 JOURNAL MEDLINE PUBMED  
 TITLE COMMENT  
 CONTACT: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,M., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Kono,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watanabe,A., Yamashita,M. and Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Exploration Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
 FEATURES  
 SOURCE location/Qualifiers  
 1. 439  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

BASE COUNT 97 a 116 c 122 g 104 t  
 ORIGIN  
 Query Match 66.7% Score 22; DB 13; Length 439;  
 Best Local Similarity 78.1% Pred. NO. 5.8e+02;  
 Matches 25; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 1 CCTGCAACTGCTGCTCTCCAGCATGACG 32  
 78 CCTGCAACTGCTGCTCTCCAGCATGACG 47  
 RESULT 12  
 LOCUS BY535446/C  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 REFERENCE 1 (bases 1 to 463)  
 AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nishida,I., Osato,N., Saito,R., Suzuki,H., Yamana,H., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bull,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Brad,D., Brusch,V., Chotila,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimond,S., Guernicich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltas,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perle,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Saitana,R., Takenaka,Y., Taylor,M.S., Teasdale,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,M., Tomita,M., Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavalan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Kono,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shingawa,A., Yasuniishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.  
 ANALYSIS: The mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 NATURE 420, 563-573 (2002)  
 22354683  
 12466851  
 JOURNAL MEDLINE PUBMED  
 TITLE COMMENT  
 CONTACT: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,M., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Kono,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watanabe,A., Yamashita,M. and Hayashizaki,Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Exploration Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
 FEATURES  
 SOURCE location/Qualifiers  
 1. 463  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Substantiation  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedic: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

# FEATURES

## source

1.463 location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="NOD"  
/db\_xref="taxon:10090"  
/clone="P630305G21"  
/cell\_type="NOD-derived CD11c +ve dendritic cells"  
/clone\_lib="RIKEN full-length enriched, NOD-derived CD11c +ve dendritic cells"  
BASE COUNT 102 a 120 c 131 g 110 t

Query Match 66.7%; Score 22; DB 13; Length 463;  
Best Local Similarity 78.1%; Pred. No. 5.9e+02;  
Matches 25; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTGCAACTCGTGTCTCCAGCATGACGAG 32  
Db 96 CCTGCCATCTGTGTCTCTGCGCAGACAG 65

RESULT 13 B1528489 643 bp mRNA linear EST 29-AUG-2001  
LOCUS 1024091A09.y1 C. reinhardtii CC-1690, normalized, lambda Zap II  
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION B1528489  
VERSION B1528489.1 GI:15369063  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonads.  
1 (bases 1 to 643)  
REFERENCE Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1024b  
JOURNAL Unpublished  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159

Fax: 919 613 8177  
Email: [chauser@duke.edu](mailto:chauser@duke.edu)  
location/Qualifiers  
1.463  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, lambda Zap II"

# FEATURES

## source

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldi et al (1996) Genome Research 6: 791-806."  
BASE COUNT 126 a 178 c 226 g 113 t

Query Match 66.7%; Score 22; DB 12; Length 643;  
Best Local Similarity 78.1%; Pred. No. 6.5e+02;  
Matches 25; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 2 CCTGCAACTCGTGTCTCCAGCATGACGAG 33  
Db 112 CATTGCAACCGTGTCTCTCAGCATGACGAG 143

RESULT 14 B1528424 651 bp mRNA linear EST 29-AUG-2001  
LOCUS 1024090E05.y1 C. reinhardtii CC-1690, normalized, lambda Zap II  
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION B1528424  
VERSION B1528424.1 GI:15368998  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonads.  
1 (bases 1 to 651)  
REFERENCE Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1024b  
JOURNAL Unpublished  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: [chauser@duke.edu](mailto:chauser@duke.edu)

# FEATURES

## source

1.461 location/Qualifiers  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, normalized, lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the

light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO<sub>2</sub> and HS medium bubbled with 5% CO<sub>2</sub>. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with Exsistat (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 127 a 179 c 230 g 113 t 2 others

ORIGIN

Query Match 66.7%; Score 22; DB 12; Length 651;  
Best Local Similarity 78.1%; Pred. No. 6.5e+02;  
Matches 25; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 CCTGCAAACTGCTGCTCTCCAGCATGACGG 33  
DB 112 CGTGCACACCCGTTGCTCTCCATGCAGGG 143

RESULT 15  
AO587305/c 398 bp DNA linear GSS 07-JUN-1999  
LOCUS RPCI-11-452E3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-452E3,  
DEFINITION genomic survey sequence.

ACCESSION AO587305

VERSION AO587305.1 GI:5013985

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 398)

Zhao,S., Adams,M.D., Nieman,W., Malek,J., de Jong,P. and Venter  
,J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

Map Building

Unpublished

Other GSSes: RPCI-11-452E3.TV

Contact: Shaying Zhao, William Nieman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html.

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 398

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="GDB:7673282"

/db\_xref="taxon:9606"

/clone="RPCI-11-452E3"

/sex="Male"

/cell\_type="Lymphocytes"

/clone\_lib="RPCI-11"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPCI11 Human Male BAC Library"

BASE COUNT 73 a 117 c 98 g 101 t 9 others

ORIGIN

Query Match 66.1%; Score 21.8; DB 28; Length 398;

Best Local Similarity 82.1%; Pred. No. 6.7e+02;

Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 CCTGCAAACTGCTGCTCTCCAGCATG 28  
DB 104 CCTGCAACCCCGTTGCTCTCCATGCAG 77

Search completed: September 16, 2003, 20:40:01  
Job time: 1674.26 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:35:29 ; Search time 657.921 Seconds

(without alignments)  
1865.405 Million cell updates/sec

Title: US-09-594-065-3

Perfect score: 30

Sequence: 1 cccctgcgtagtcgtacgacctccgcaggg 30

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

```

1: GenBank:
2: gb_ba:
3: gb_hcg:
4: gb_in:
5: gb_om:
6: gb_ov:
7: gb_pac:
8: gb_ph:
9: gb_pl:
10: gb_pr:
11: gb_ro:
12: gb_sfr:
13: gb_by:
14: gb_un:
15: gb_vl:
16: em_ba:
17: em_fun:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pac:
24: em_ph:
25: em_pl:
26: em_ro:
27: em_sfr:
28: em_un:
29: em_vl:
30: em_hcg_hum:
31: em_hcg_inv:
32: em_hcg_other:
33: em_hcg_mus:
34: em_hcg_dln:
35: em_hcg_rtd:
36: em_hcg_mam:
37: em_hcg_vrt:
38: em_hcg:
39: em_hcg_hum:
40: em_hcg_mus:
41: em_hcg_other:

```

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.6	68.7	40281	1	MLCB22
2	20.6	68.7	348450	1	MLCB22/c
3	20.4	68.0	11282	1	AB003979
4	20.4	68.0	302835	1	AB012555
5	20.2	67.3	110000	2	LMFLCHR34_04
6	19.4	64.7	68974	2	AC074063
7	19.4	64.7	110079	10	AC002315
8	19.4	64.7	163132	2	AC121312
9	19.4	64.7	176222	2	AC117138
10	19.4	64.7	194985	10	AC002406
11	19.4	64.7	200548	10	AL672026
12	19.4	64.7	217622	2	AC126569
13	19.4	64.7	222869	2	AC113777
14	19.4	64.7	226382	2	BX294168
15	19.4	64.7	226790	2	AC107570
16	19.4	64.7	229040	2	BX510361
17	19.4	64.7	231268	2	AC118318
18	19.4	64.7	241998	2	AC112739
19	19.4	64.7	263091	2	AC120822
20	19.4	64.7	267971	2	AC128995
21	19.4	64.7	270706	2	AC129862
22	19.4	64.7	299800	1	AP005040
23	19.2	64.0	182462	9	AC092669
24	19.2	63.3	94255	9	AC003090
25	19.2	63.3	110000	2	AC15691_2
26	19.2	63.3	122279	2	AC128646
27	19.2	63.3	146432	2	AC079358
28	19.2	63.3	166860	2	AL451002
29	19.2	63.3	182459	2	AC116818
30	19.2	63.3	197448	2	AC138342
31	19.2	63.3	200321	2	AC129934
32	19.2	63.3	207265	10	AC125157
33	19.2	63.3	211397	10	AC098887
34	19.2	62.7	576	6	AX122231
35	18.8	62.7	576	6	BD164348
36	18.8	62.7	702	9	HSB340406
37	18.8	62.7	732	9	HSB339685
38	18.8	62.7	751	9	HSB340046
39	18.8	62.7	753	9	HSB340037
40	18.8	62.7	793	9	HSB340033
41	18.8	62.7	92019	9	AL450320
42	18.8	62.7	94244	9	AC063938
43	18.8	62.7	110000	2	AC003656_6
44	18.8	62.7	113454	2	AC017944
45	18.8	62.7	120871	9	AP001054

## ALIGNMENTS

RESULT 1  
LOCUS MLCB22/c  
DEFINITION Mycobacterium leprae cosmid B22.  
ACCESSION Z98741  
VERSION Z98741.1 GI:2342602  
KEYWORDS 30S ribosomal protein S15; ABC-type transporter; aminomethyl transferase; aminopeptidase; branched-chain amino acid aminotransferase; cobT; cybB; cybK; cysteine synthase; dihydrolipoamide succinyltransferase; gcvT; gpeI; guanosine pentaphosphate synthetase; llyE; llyB; llyP; lipote-protein ligase; lipoic acid synthetase; major membrane protein I; mmpI; multifunctional enzyme; nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase; nifs-like protein; oxidoreductase; pepA; polyribonucleotide

SOURCE  
 ORGANISM  
 Mycobacterium leprae  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacteriineae; Mycobacteriaceae; Mycobacterium.  
 REFERENCE  
 AUTHORS  
 TITLE  
 1 (bases 1 to 40281)  
 Biglieri, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.  
 Use of an ordered cosmid library to deduce the genomic organization  
 of Mycobacterium leprae  
 JOURNAL  
 MEDLINE  
 PUBMED  
 93188700  
 8446027  
 2 (bases 1 to 40281)  
 Devlin, K. and Churcher, C.M.  
 REFERENCE  
 AUTHORS  
 TITLE  
 3 (bases 1 to 40281)  
 Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
 Direct Submission  
 JOURNAL  
 Submitted (22-AUG-1997) Mycobacterium leprae sequencing project,  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr.  
 Stewart T. Cole, (3) Unite de Genetique Moleculaire Bacterienne,  
 Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15,  
 France Requests for cosmids should be sent to Karin Biglieri  
 (kebig@pasteur.fr)  
 COMMENT  
 Notes:  
 The Sanger Centre is funded to complete the sequence of M. leprae  
 by the Helser Program for Research in Leprosy and Tuberculosis of  
 The New York Community Trust.  
 Work in Paris is supported by the Helser Trust, the Association  
 Francaise Raoul Follereau and the Groupement de Recherches et des  
 Etudes des Genomes (GEP-GRG).  
 Details of M. leprae sequencing at the Sanger Centre are available  
 on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/>)  
 CDS are numbered using the following system eg MLCB33.01c. ML (M.  
 leprae), c333 (cosmid name), .01 (first CDS), c (complementary  
 strand).  
 The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous. The length  
 in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given  
 where these have been used to deduce the initiation codon. All CDS  
 over 100 codons have been analysed. Gene prediction is based on  
 positional base preference in codons especially where there is an  
 increase in the observed/expected third position G + C. CAUTION:  
 We may not have predicted the correct initiation codon. Where  
 possible we choose an initiation codon (atg, gtg, or ttg) which is  
 preceded by an upstream ribosome binding site sequence (optimally  
 5-13bp before the initiation codon). If this cannot be identified  
 we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the  
 sequenced clone. It may be shorter because we only sequence  
 overlapping sections once, or longer, because we arrange for a  
 small overlap between neighbouring submissions. Cosmid B22 is  
 overlapped by B151 at the 5' end, and by B1229 at the 3' end.  
 There are at least 9 conflicts between this sequence and the  
 previously published sequences; in each case our sequence has been  
 checked and is thought to be correct.  
 FEATURES  
 SOURCE  
 1..40281  
 /organism="Mycobacterium leprae"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1769"  
 /clone="cosmid B22"  
 complement(1..7427)  
 /note="overlap with EM\_BA\_MLB151CS L78813 Mycobacterium  
 leprae cosmid B1551 from: 2 to: 7425"  
 complement(1..198)  
 misc\_feature  
 gene

CDS  
 /gene="MLCB22.01c"  
 complement(1..198)  
 /gene="MLCB22.01c"  
 /note="MLCB22.01c, unknown, partial CDS, len >66 aa;  
 similar to hypothetical proteins from a range of organisms  
 eg. YADR\_ECOLI\_P37026 hypothetical 12.1 kd protein in  
 hemA-pfe intergenic region (114 aa), fasta scores: Opt:  
 164 z-score: 323.9 E(): 6.3e-11, 45.1% identity in 51 aa  
 overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein MLCB22.01c"  
 /protein\_id="CAB11366.1"  
 /db\_xref="GI:2342603"  
 /db\_xref="SPTREMBL:O32951"  
 /translation="MAVQNETSAKHGVILTDVATKAKSLLDGSRDMLRIAVQP  
 GCGAGLRNLPFDRLTDGDLT"  
 complement(373..567)  
 /note="MLCB22.02c, probable pseudogene fragment, len 64  
 aa; similar to part of Y00G MYCTU Q10394 hypothetical 48.6  
 kd protein cy190.16 (480 aa) (57.8% identity in 64 aa  
 overlap)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /db\_xref="PID:e335004"  
 /db\_xref="PSEUDO:CAB11367.1"  
 complement(566..670)  
 /note="MLCB22.03c, probable pseudogene fragment, len: 34  
 aa; similar to part of Y00G MYCTU Q10394 hypothetical 48.6  
 kd protein cy190.16 (480 aa) (76.5% identity in 34 aa  
 overlap)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /db\_xref="PID:e335005"  
 /db\_xref="PSEUDO:CAB11368.1"  
 complement(799..861)  
 /note="MLCB22.04c, probable pseudogene fragment, len: 20  
 aa; similar to part of Y00G MYCTU Q10394 hypothetical 48.6  
 kd protein cy190.16 (480 aa) (75.0% identity in 20 aa  
 overlap)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /db\_xref="PID:e335006"  
 /db\_xref="PSEUDO:CAB11369.1"  
 complement(1092..1181)  
 /note="MLCB22.05c, probable pseudogene fragment, len: 29  
 aa; similar to part of Y00G MYCTU Q10394 hypothetical 48.6  
 kd protein cy190.16 (480 aa) (63.0% identity in 27 aa  
 overlap)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /db\_xref="PID:e335007"  
 /db\_xref="PSEUDO:CAB11370.1"  
 complement(1220..1312)  
 /note="MLCB22.06c, probable pseudogene fragment, len: 30  
 aa; similar to part of Y00G MYCTU Q10394 hypothetical 48.6  
 kd protein cy190.16 (480 aa) (63.3% identity in 30 aa  
 overlap)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /db\_xref="PID:e335008"  
 /db\_xref="PSEUDO:CAB11371.1"  
 1478..2221  
 /gene="MLCB22.07"  
 1478..2221  
 /gene="MLCB22.07"  
 /note="MLCB22.07, unknown, len: 247 aa; highly similar in  
 C-terminus to M. tuberculosis Y00H\_MYCTU Q10395 very

```

JOURNAL Nature 409 (6823), 1007-1011 (2001)
MEDLINE 2128732
PUBMED 11234002
REFERENCE 2 (bases 1 to 348450)
AUTHORS Parkhill,J.
TITLE Direct Submmission
JOURNAL Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
Centre, Hinxton, Cambridge, CB10 1SA, UK Unité de Génétique
Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux,
75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
COMMENT Notes:
Details of M. leprae sequencing at the Sanger Centre are available
from http://www.sanger.ac.uk/projects/M\_leprae/. A relational
database containing the M. leprae sequences is available from
http://genolist.pasteur.fr/Leproma/.
location/Qualifiers
FEATURES
source 1..348450
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/script="TN"
/db_xref="taxon:1769"
551..1786
/gene="ML0842"
551..1786
/gene="ML0842"
551..1786
/gene="ML0842"
/note="Similar to Mycobacterium tuberculosis hypothetical
44.6 kDa protein Rv1464 or MT007.11 TR:O53155
(EMBL:AL021184) (417 aa) fasta scores: E(): 0, 43.6% id in
408 aa and to many other bacterial nifs-homologues, e.g.
Bacillus subtilis YnrW protein ynrW TR:O32164
(EMBL:D29120) (406 aa) fasta scores: E(): 0, 46.7% id in
405 aa. Previously sequenced as TR:O32975 (EMBL:Z58741).
Contains Pfam match to entry PF00266 aminotran_5,
Aminotransferases class-V.
Similar to ML0117, ML0596 and ML1708"
/codon_start=1
/transal_table=11
/product="conserved hypothetical protein"
/protein_id="CAC31223.1"
/db_xref="GI:13092923"
/transalation="MTDVPDIQAVRADPILOSTVNGKPLIMPNAATQKQPVIVDR
LSFPYHENSNTIRAAHELRATDAVERARETARRRTGAKAOEITFRKTTAININ
VAAWGGKHLPQGEDEVVITLERRANIPVMQLSGRGARVLKAPVDAGNLMSFEF
DLGPRTKILVAATOVNSALGTVTQGEKIVLGHSYGARVIDGASIPHPINSEIG
ADPFPSGHKIYPTGTGIVLGCEDVTEPMPOGSGNMIVDTLESLVQGPNNKEEF
AGCNFLADNAGTGALRYRVCVORIASHGALLDYATPRADIIPVRVLGVTETKRKA
SVSPFLAGHPELVGKALNABGIAVRAGHCQAQVLRRLGLEATVTRPSAFNYTTEE
IDVFINVRIARGTGNIC"
653..1738
/gene="ML0842"
/note="Pfam match to entry PF00266 aminotran 5,
Aminotransferases class-V, score 23.70, E-value 1.7e-08"
1908..2395
/gene="ML0843"
/note="synonym: lppP"
/pseudo
1908..2395
/gene="ML0843"
/note="synonym: lppP"
/pseudo
1908..2395
/gene="ML0843"
/note="Possible pseudogene of M. tuberculosis orthologue
lppP (Best blastx score 294)"
/pseudo
/codon_start=1
/transal_table=11
/product="lipoprotein (pseudogene)"
2349..2555
/note="possible RBS"
2559..4112
/gene="ML0844"
/note="synonym: narK"
2559..4112
/gene="ML0844"
/note="Similar to Mycobacterium tuberculosis putative

```

nitrite extrusion protein narX1 or Rv2329c or MTCY312.05  
 TR:P71883 (EMBL:Z79702) (515 aa) fasta scores: E(): 0,  
 69.3% id in 488 aa and to Escherichia coli nitrite  
 extrusion protein 2 narX SW:NRU\_ECOLI (P37558; P7696)  
 (462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There  
 is a frameshift near the C-terminus relative to the M.  
 tuberculosis homologue. Previously sequenced as TR:032974  
 (EMBL:Z98741). Contains hydrophobic, probable  
 membrane-spanning regions."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative nitrite extrusion protein"  
 /protein\_id="CAC31225.1"  
 /db\_xref="GI:13092924"  
 /translation="MDQVYLQNEBSLYGGRSKTPIDHRSLKSSPFRRLRAC  
 RISHMDPEQDAWEAGNKTIALNLMSVLTTHLGQSVWTLPMWELMPKQVPSA  
 GKFLATTATLVGACLRVPSLATLTFEGRMWALPSWVLLPTITAVLAAHGLP  
 LWPYLCALTLGSGNFPASMTNNAFYPHRLKGAFLAGAGNLGVSIVQWGRD  
 VIASVDRKPYIVCGLYVLLIAGIAMLFNNDIEHRIQVITRPVLFVVSTRDS  
 NVIALIYLASFGSFGSPAGCVLETNVPAGOSTPAQALAAELAFGLTAAVAR  
 FPGRLADRLGSRVTLVVGAMVFAAGLGLVGLIEGSRVCPRIQVMAVYFAGIT  
 LFGIGANGSVYKMTPTIPKASHSLGINDCECDMSRVISGVYGFAYAGAGV  
 GTDLARESYLNTGVTAAFWIPMLCTAAGVLTWMTVCPLPGLNDEANNAFAS  
 VASRTHRG"  
 /complement(4100..4681)  
 /gene="ML0845"  
 /complement(4100..4681)  
 /gene="ML0845"  
 /note="Unknown function. Similar to part of some acyl-CoA  
 oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2  
 TR:065201 (EMBL:AF057043) (692 aa) fasta scores: E():  
 1.5e-06, 35.2% id in 125 aa. Previously sequenced as  
 TR:032973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0,  
 99.5% id in 193 aa."  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAC31226.1"  
 /db\_xref="GI:13092925"  
 /translation="MELRPTNAAASERVTRKATETTLQRMVGAROSEKKSGLFNRGT  
 OTIPKCECDLSAVARRQYKSKEMRAPPTNPYQDHLAAEHTRVYLFETFLA  
 GIESCDNPARKLIDVCDLVASVARKGTSSIDSPSALKAIAKSGPMPMSAAL  
 CRDVGRFSESSQSCVTPRCRTSRICPSLQPR"  
 /complement(4959..5480)  
 /gene="ML0846"  
 /pseudo  
 /complement(4959..5480)  
 /gene="ML0846"  
 /note="Possible pseudogene similar to M. tuberculosis  
 paralogue Rv1747 (best blastx score 127)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter (pseudogene)"  
 /complement(5522..5848)  
 /gene="ML0847"  
 /note="Possible pseudogene of M. tuberculosis orthologue  
 Rv2327 (best blastx score 184)"  
 /pseudo  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein (pseudogene)"  
 /complement(5906..8080)  
 /gene="ML0848"  
 /complement(5906..8080)  
 /gene="ML0848"  
 /note="Similar to Mycobacterium tuberculosis hypothetical  
 ABC transporter ATP-binding protein Rv2326c or MTCY312.08  
 SW:YN26\_MYCTU (P71886) (697 aa) fasta scores: E(): 0,  
 76.9% id in 697 aa. Shares similar domains with many

ABC-type transporters e.g. Streptomyces roseofulvus AtPase  
 component of putative ABC transporter kind TR:068910  
 (EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5%  
 id in 495 aa and Synchococcus sp. nitrate transport  
 ATP-binding protein NrtD nrtD SW:NRTD\_SYNP7 (P38046) (274  
 aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa.  
 Previously sequenced as TR:032971 (EMBL:Z98741). Contains  
 hydrophobic, possible membrane-spanning regions. Contains  
 2 Pfam matches to entry PF00005 ABC tran, ABC transporter.  
 Contains 2 x P800017 ATP/GTP-binding site motif A  
 (P-loop). Contains 2 x P800211 ABC transporters family  
 signature."  
 /codon\_start=1  
 /transl\_table=11  
 /product="ABC transporter"  
 /protein\_id="CGC31229.1"  
 /db\_xref="GI:13092926"  
 /translation="MDLCIHRMALLCOATRCRHKQTQLGCAADLPAGTGTGKV  
 TADHRPHLHRSGLQPELQAQSVLAACVTAIVSVVPPAAGALIGVPMELLA  
 YRFRVMTNMAVAGVAPLITIGGFIAVNSAYIGLGVGRKCGTLYEAL  
 LPAGLAFGAANVAVLVIGRLRLFKAMTANVDIAATLPMELPMVAVOLKRYEAD  
 LGQHPMMLIGYFVITLIVSLIGMWVLSVLERTRDIPDHKLDAPEACNDAVGP  
 VFWLIDKRFRRYPHAGDALARBSVDLAEVGEHVAVTANGSKTTMLIAGREPTSG  
 TVDRGAVGLGSLGTAVALDHPESQVLTAVADVNGLPETVDVNRLLREYGLD  
 AFABERDTGSLSGELQRLAALAAAREBSLLIADSVTSMVROGDALIGVSLTKR  
 HPIALVHTHYNNEADTADRTINLSDSDPNMGSETVAPVSTVADHPPHAPLELV  
 GVGEYGSGETPMAKALADHSFVNVGGGVYVYNGSGSKTIAIMHAPVETGAC  
 LIDGPTHEHGAVALSFQAAIOLMSRVRDLEVASAGFSPDRDRAVAAAGVGLD  
 PALAKRIDQSGCGMRVYVLAGLIQCSRPALLIDEPRLAGIDAVSQRLRLLEDLRC  
 EKGITVVTSHDPSGLDVCPTVRLNGLAESVITAGTS"  
 /gene="ML0848"  
 /note="Pfam match to entry PF00005 ABC tran, ABC  
 transporter, score 112.30, E-value 9e-30"

Query Match 68.7%; Score 20.6; DB 1; Length 348450;  
 Best Local Similarity 85.2%; Pred. No. 1.8e+02;  
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCTGGTAGTGTACGACCTCCTGCAG 28  
 Db 5076 CTTAAGTAGTGCAAGACCTCCTGCAG 5102

RESULT 3  
 AEO03979  
 LOCUS 11282 bp DNA linear BCT 15-JUN-2001  
 DEFINITION Xylella fastidiosa 945c, section 125 of 229 of the complete genome.  
 ACCESSION AEO03979 AEO03849  
 VERSION AEO03979.1 GI:9106520  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Xylella fastidiosa 945c  
 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 Xanthomonadaceae; Xylella.  
 1 (bases 1 to 11282)  
 REFERENCE  
 AUTHORS  
 Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,  
 Alvarenga,R., Alves,L.M., Araya,J.E., Bata,G.S., Baptista,C.S.,  
 Barros,M.H., Bonaccorsi,E.D., Borlin,S., Bove,O.M., Brites,M.R.,  
 Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrier,H.,  
 Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,  
 Coutinho,L.P., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,  
 Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,  
 Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Partlan,L.R.,  
 Garneri,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,  
 Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,B.L., Kitajima,J.P.  
 and Marino,C.L.  
 TITLE  
 The genome sequence of the plant pathogen Xylella fastidiosa. The  
 Xylella fastidiosa Consortium of the Organization for Nucleotide  
 Sequencing and Analysis  
 JOURNAL NATURE 406 (6792), 151-157 (2000)  
 MEDLINE 20365717  
 PUBMED 10910347



gene

CDS

gene

CDS

```

/gene="XF1498"
/complement(7506..9266)
/gene="XF1498"
/note="similar to SP|P52673 (percent identity: 52 %/query
alignment coverage: 93.0 %/subject alignment coverage:
97.5 %) : identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/cranisl_table=1
/product="NADPH-sulfite reductase, iron-sulfur protein"
/protein_id="AAF84307.1"
/db_xref="GI:9106527"
/translation="MSPTSLIILHPGSPAPMSSESDIKGSHRLRGSLIESLN
PTGLAHSDDTLIKTHGSYOQDDRDAPERRRQKLEPAQPMIRTRTPGCVTPOOM
LQDLAIRVYNHSLRVTROAFQPHGVTKEELATQALNALITDLACGVNKNV
OVANPKISRAHADLYTDAALSHLLPNTAYVEIWLDRKQVAGAEEREPYGPYH
LPRAKKIFPAAPINDVDVFANDGLFIAVDNTLGVNAILGGMGCTTGDDTPMR
VNTIIGFTIRADLAIALSTAIYTTORNGNTRLRKARPKYTIIDRGGLDCLVGRILQORA
GITLOPAPRPVFEHNGDRGVWIEEGEDRHHMTLSLPAGRCIADTRESSPLSGPAINQL
GITQGRMTPNQNVIVAGISPGQARITDLVTQYDLOGNGAPATRLAAHNAACVALPFC
GLAMAEARVYLPDFPVNKLQPILEKRYGLAEKRIILRISGCGNGSGRPVLAIALVGAP
GSVYNMLGGDGRGRLNLTLYRENTETETIILAAEPLILGRVOOKRLPGEFGDPLHRTG
IIALPPYPTHHRVISSTLOA"
/complement(9336..11174)
/gene="XF1499"
/complement(9336..11174)
/gene="XF1499"

```

Query Match	68.0%	Score 20.4	DB 1	Length 11282
Best Local Similarity	80.0%	Pred. No. 1.7e+02		
Matches 24; Conservative	0	Mismatches 6	Indels 0	Gaps 0

```

Oy      1  CCCTGCGTAGTGTGACGACCTCCTGCAGGG  30
          |||||  |||  |||||  |||||
Db      7107 CCCTGCGCATCGGTTGCACCTTGTGCAGGG  7136

```

RESULT 4	AE012555	LOCUS	DEFINITION
	AE012555	302635 bp	DNA linear BCT 29-JAN-2003
			<i>Xylella fastidiosa</i> Temecucal, section 3 of 9 of the complete

ACCESSION	AE012555	AE009442
VERSION	AE012555.1	GI.28056497
KEYWORDS		
SOURCE		
ORGANISM	<i>Xylella fastidiosa</i>	<i>Temeculana</i>
	<i>Xylella fastidiosa</i>	<i>Temeculana</i>

REFERENCE  
AUTHORS  
1 (bases 1 to 302835)  
Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B.,

Civerolo, E. L., Simpson, A. J. G., Almeida Jr., N. F., Setubal, J. C. and Kitzajima, J. P.

J. Bacteriol. 185 (3), 1018-1026 (2003)

## REFERENCES

J. Bacteriol. 185 (3), 1018-1026 (2003)  
12533478  
2 (base 1 to 302835)  
Van Sluys, M.A., de Oliveira, M.C., Monteiro-Vitorello, C.B.,  
Miyaki, C.Y., Purlan, L.R., Camargo, L.E.A., da Silva, A.C.R.,  
Moon, D.H., Takita, M.A., Lemos, E.G.M., Machado, M.A., Ferro, M.I.T.,

TITLE	Direct Submission
JOURNAL	Submitted (17-DEC-2001) Instituto de Biociencias, Universidade de
FEATURES	Sao Paulo, Rua do Matoso, 277, Sao Paulo, SP 05508-900, Brazil
SOURCE	Location/Qualifiers
	1. .302835

gene  
CDS

```
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
```

```

/protein_id="AA028377.1"
/db_xref="GI:28056498"
/translation="MVGMSRIILNNLFFP

```

EKA VMDRA YPCGNFPRRTIEGALTLTSPDAOCLRRRMBLIDLGJNMRITQPGGASTDPL  
NTEQGPYPSRSHPLNSNDRGVYEVSVRKKVCDASVJESNCVAYPSGTYAPRGCJVOEY  
SKRVSYSFVSKNDYLLDDGCVLRAROKPVGPGTYTPEQCEKPLPAHEMIRPGVLYVY  
DNPNEDEKATLTHRVGRTIGNSGVINTLNKRSQMETGKNTKGPVSELYYTAIRYPRR

PENTNTTGEPRKPOAVNRDRSIDVVKMAQJFQHEGMRQODAMTSALPSNPLVPGA  
GNSNYSIALYADATKDMRDLGDDQLLTHHWDVVEGGDYKJPISTNOYMLAKYG  
GFQVAGYDPEDKTNPPLSEATYNTNGEIVNGDPTAKADNDIYAADEXVYALIKAR  
SIVAEIAGAGTGLSSNARLETGAVTQAFSGTARGDILAHVOKVYCALPLPNWAL  
ANFPMERDVUKENATTLQDPTKKNIGQTLASASQOINYLGDSSQEBNVEGKLR

VTGLGRGGKGLFALDVTDPSPNIRLLMDKTSADIGGLGNTLSKPMIAQTSDDGTWSVLGG  
NGPNSTADNAQLIMNLLTGHAAQVPVSKTSNNGLSGVFPMSQSNGLTIDRVYAGDLL

```
/codon_start=1
/transl_table=11
```

```

/translation="MLRFATAGFFVYLLTPMVKASTPHYSVNTPFNQSTPTTVTPIN

```

gene

```

/translucation="MRPATAGGFVLLTPMVKASTHYSYNTFPNOSTPPTYVTPIN
TMFNPQPPPLGPGKALSORYNRELYMDPMLHLRVNSROSIAKTSTSTSTANSNN
KTEIYDKQTYDYNTPMRFNNQCKRTADEFAMMKARGVHIPOSTNNNVQSVBBA
AK"
complement(5836..6225)
 locus_tag="PD0504"
```

CDS  
complement(5836..6225)  
/locus\_tag="PD0504"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA028379.1"  
/db\_xref="GI:28056500"  
/translation="MNKFTAMTRALRKHMTDLAQOANDRIRRTAPQASOMLKSAALE  
VKSRRKVTGKLIRKNSPIAVAVSGLAGYMTYRKYKMKKKKIYINGGEMLN  
EHADHYEVRBSIHRHRSAAATDNNNG"  
complement(6491..9001)  
/gene="hrpB"  
/locus\_tag="PD0505"  
complement(6491..9001)  
/gene="hrpB"  
/locus\_tag="PD0505"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent helicase"  
/protein\_id="AA028380.1"  
/db\_xref="GI:28056501"  
/translation="MCRMNHALPIDALLPAITTSLSNHPRLILBAPPAGKTTQVPL  
ALLDAPMLNGRSIIILBPRVAVAAALFMAOCGEAGLGVYHIFRESKSAARTI  
HYVTEGLITRIIODDPLDLSIGALFDEPHERHLADGIALALDVOVLPEDLITVY  
MSATIQGRIARLPALRLTSGSRAYPTTITFPAPRBSLETQSRALIQHALDTHOG  
DVLITLPQORETARLQALILEKTLSPELHPLHGLPLQEBOTRVLEPDQGRRLVLA  
TNVASSITLPGISVIDSGAREPAPYDPTNGFTLDVASTIQASADQAGARLAP  
GMAYRIEMPQSORLEAORPELIQTELSGALLETAMGSSITLAFIDPPSGALAREL  
LORDGALISNTSTLPLGKMLAIGHPLAAVLAPHDPOEALACDLALLDAREPL  
RSGSGALISRMALAEKRSGLPHDTKSTLKALDATHQMRIRCDATPDPAIEAH  
TLGNLTITAFPERIARHSDDPLVYLANGRMRLDLSALGEMWITASELHREKID  
IOLUCGAPDERKHLADFPERTIQHVIKMRDOKRALVAQRSCFDRIYLDVPRGRI  
DVOQANVLTGAVRTGLDITLSTWELRQWQARAVALLHMIPELPLSNTTLATL  
DTWLPAFAGKTRLDALDEAFAEQAVKSLSTQHALIDRHVPAHIIVPSGGRIDY  
GLADGTPQPPVLSYQBLFGLPTPRIAKNRIPLTTLHSPGRPLQVTDLSFV  
TNTYEVKEMKGRYPKHPMPDDPMTATATHTKSPGK"  
complement(9502..9990)  
/locus\_tag="PD0506"  
complement(9502..9990)  
/locus\_tag="PD0506"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA028381.1"  
/db\_xref="GI:28056502"  
/translation="MLMLAPFLTNGMKPROVAQLEAQAALIHGETRIANAANLAL  
IFHSLPLANWGFYTDGRELVEYFGQIPACIRIPLKGVCGAARQTORITDVID  
TFPHIACDSASRSELVPLKNDTLIGVLIDDSPTCARFSDQDTGIESLARVSN  
LR"  
10282..12135  
/locus\_tag="PD0507"  
10282..12135  
/locus\_tag="PD0507"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="pathogenicity protein"  
/protein\_id="AA028382.1"  
/db\_xref="GI:28056503"  
/translation="MEVLRCFNVNRFITLSILPRYTAQMLRRFIPMLSLWLFASLA  
QANATVDKIQIGLKDGVCEMMKNITVSLSYVSGKEGSESLLEYLSQEPQTRQ  
ALEFGYAPAIRIDAPRONDHITVIVYDKEPVRIRQAHVMAAGAAADHYLORDI  
EDPKLGEIRIHPITPEASKYRITPRLAEGYADPFRROVYEVTRATYAADIDIME  
SGRYDMGPRFHTIDYFHEGLFNPVLTBEGSYFHEGLDRKESLTIKDIFSSIDIO  
KPEEADSEGNVAVKLERAKSKITYGISTYSSSGALRAGVERRYMAAGHMA  
RLDYAONKSLITTAOIPAPKWLDMYIFSAAYDEQOYIDLKRVKLSAASGQINR

gene  
CDS  
HLTATASINALREBRMYAADGTNTVAYQHSITLVYPOLEASYVDVDAATFPRNGSAAT  
VILRGASAFGSKSNFTQLHQLMFMHGLGSSRLILGEGTTWASDLVAMPPLIRF  
FAGVNSIRGAYAPREVGRRTAKPDPAFALGAGHVSAGAEVYVYXGCPFGGAVFVDSG  
SAFNRYPMDHTGIGIGLRYRSPVGPVRVDIARGLNKPSSKIQLYIDIANL"  
12414..15953  
/locus\_tag="PD0508"  
12414..15953  
/locus\_tag="PD0508"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="pathogenicity protein"  
/protein\_id="AA028383.1"  
/db\_xref="GI:28056504"  
/translation="MDPIMRLILRLRLDVLKMSNASTLVSQSEPEPKLPSPMDVLP  
SINYPALQITITITIDNRISIKQPLIHLRIKAVNAVAGSELIQALYDSRGN  
RADYTRPRHDYSDLTATVAVPSVRGPPASFGVIGDLDRLQALAGRAPVLA  
DVNLNGRSDPMTFSASAEHPFGLLPGELTGPPTAPALA FNYHAAQOGKADLKQ  
FKQGEQITVLQPSRLQLANQMLNWEPLVIDAFGGRTNLRGVDIRDPNTRPHFSVNA

Query Match 68.0% Score 20.4; DB 1; Length 302835;  
Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
1 CCTGCGTAGTGTACGACCTCTGCAGG 30  
||||| ||||| ||||| ||||| |||||  
DB 276037 CCTGCGTAGTGTACCTTGCAGG 276066

RESULT 5  
LMFLCHR34\_04  
WCOMMENT"  
Sequence split into 18 fragments  
Fragment Name Begin End  
LMFLCHR34\_01 1 110000  
LMFLCHR34\_02 100001 210000  
LMFLCHR34\_03 200001 310000  
LMFLCHR34\_04 300001 410000  
LMFLCHR34\_05 400001 510000  
LMFLCHR34\_06 500001 610000  
LMFLCHR34\_07 600001 710000  
LMFLCHR34\_08 700001 810000  
LMFLCHR34\_09 800001 910000  
LMFLCHR34\_10 900001 1010000  
LMFLCHR34\_11 100001 1110000  
LMFLCHR34\_12 1100001 1210000  
LMFLCHR34\_13 1200001 1310000  
LMFLCHR34\_14 1300001 1410000  
LMFLCHR34\_15 1400001 1510000  
LMFLCHR34\_16 1500001 1610000  
LMFLCHR34\_17 1600001 1710000  
LMFLCHR34\_18 1700001 1810000  
Continuation (5 of 18) of LMFLCHR34 from base 400001 (AL499623 Leishmania major chromosome  
Query Match 67.3% Score 20.2; DB 2; Length 110000;  
Best Local Similarity 88.0%; Pred. No. 2.5e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
2 CCTGCGTAGTGTACGACCTCTGC 26  
||||| ||||| ||||| ||||| |||||  
DB 32750 CCTGCGTAGTGTACGACCTCTGC 32774

RESULT 6  
AC074063/c  
AC074063  
LOCUS 68974 bp DNA linear HTG 13-JUL-2000  
DEFINITION Mus musculus chromosome 18 clone RP22-32123 map 18, LOW-PASS  
SEQUENCE SAMPLING.  
ACCESSION AC074063.1 GI:9154848  
VERSION AC074063.1  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus (house mouse)

```

ORGANISM      Mus musculus
REFERENCE     Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 68974)
AUTHORS       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL       Mus musculus chromosome 18, clone RP2-32123
REFERENCE     Unpublished
AUTHORS       2 (bases 1 to 68974)
              Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Baetjen,V., Bede,F.,
              Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
              Campoliano,A., Caste,A., Choepel,Y., Colangelo,M., Collins,S.,
              Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
              Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
              Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
              Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
              Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
              Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
              McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
              Melgrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J.,
              Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
              Pisanil,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
              Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Teste,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
              Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
              Young,G., Zainoun,J., Zimmer,A. and Zody,M.
              Direct Submission
              Submitted (13-JUL-2000) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              All repeats were identified using RepeatMasker:
              Smit,A.P.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html

              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence.submissions@genome.wi.mit.edu

              ----- Project Information
              Center project name: L5920
              Center clone name: 32_1_23

              -----
              * NOTE: This record contains 81 individual
              * sequencing reads that have not been assembled into
              * contigs. Runs of N are used to separate the reads
              * and the order in which they appear is completely
              * arbitrary. Low-pass sequence sampling is useful for
              * identifying clones that may be gene-rich and allows
              * overlapping relationships among clones to be deduced.
              * However, it should not be assumed that this clone
              * will be sequenced to completion. In the event that
              * the record is updated, the accession number will
              * be preserved.
              *
              * 1
              * 741 840: contig of 740 bp in length
              * 841 840: gap of 100 bp
              * 1603 1602: contig of 762 bp in length
              * 1703 1702: gap of 100 bp
              * 2457 2456: contig of 754 bp in length
              * 2557 2556: gap of 100 bp
              * 3294 3293: contig of 738 bp in length
              * 3395 3394: gap of 100 bp
              * 4122 4121: contig of 728 bp in length
              * 4123 4122: gap of 100 bp
              * 4223 4222: gap of 100 bp
              * 4987 4986: contig of 764 bp in length
              * 5087 5086: gap of 100 bp
              * 5827 5826: contig of 740 bp in length
              * 5927 5926: gap of 100 bp
              * 6679 6678: contig of 752 bp in length
              * 6779 6778: gap of 100 bp
              * 7508 7508: contig of 730 bp in length
              * 7509 7608: gap of 100 bp

              *
              * 7609 8348: contig of 740 bp in length
              * 8349 8448: gap of 100 bp
              * 8449 9191: contig of 743 bp in length
              * 9192 9291: gap of 100 bp
              * 9292 10044: contig of 753 bp in length
              * 10045 10144: gap of 100 bp
              * 10145 10890: contig of 746 bp in length
              * 10891 10990: gap of 100 bp
              * 10991 11746: contig of 756 bp in length
              * 11747 11846: gap of 100 bp
              * 11847 12602: contig of 756 bp in length
              * 12603 12702: gap of 100 bp
              * 12703 13452: contig of 750 bp in length
              * 13453 13552: gap of 100 bp
              * 13553 14300: contig of 748 bp in length
              * 14301 14400: gap of 100 bp
              * 14401 15148: contig of 748 bp in length
              * 15149 15248: gap of 100 bp
              * 15249 16021: contig of 773 bp in length
              * 16022 16121: gap of 100 bp
              * 16122 16852: contig of 731 bp in length
              * 16853 16952: gap of 100 bp
              * 16953 17711: contig of 758 bp in length
              * 17712 17811: gap of 100 bp
              * 17812 18566: contig of 755 bp in length
              * 18567 18666: gap of 100 bp
              * 18667 19417: contig of 751 bp in length
              * 19418 19517: gap of 100 bp
              * 19518 20283: contig of 766 bp in length
              * 20284 20383: gap of 100 bp
              * 20384 21132: contig of 749 bp in length
              * 21133 21323: gap of 100 bp
              * 21323 21988: contig of 756 bp in length
              * 21989 22088: gap of 100 bp
              * 22089 22858: contig of 770 bp in length
              * 22859 22958: gap of 100 bp
              * 22959 23703: contig of 745 bp in length
              * 23704 23803: gap of 100 bp
              * 23804 24557: contig of 754 bp in length
              * 24558 24657: gap of 100 bp
              * 24658 25410: contig of 753 bp in length
              * 25411 25510: gap of 100 bp
              * 25510 26274: contig of 764 bp in length
              * 26275 26374: gap of 100 bp
              * 26374 27147: contig of 773 bp in length
              * 27148 27247: gap of 100 bp
              * 27248 28007: contig of 760 bp in length
              * 28008 28107: gap of 100 bp
              * 28107 28862: contig of 755 bp in length
              * 28863 28962: gap of 100 bp
              * 28963 29709: contig of 747 bp in length
              * 29710 29809: gap of 100 bp
              * 29810 30568: contig of 755 bp in length
              * 30569 30668: gap of 100 bp
              * 30669 31493: contig of 723 bp in length
              * 31494 31993: gap of 100 bp
              * 31993 32251: contig of 758 bp in length
              * 32252 32351: gap of 100 bp
              * 32352 33105: contig of 754 bp in length
              * 33106 33205: gap of 100 bp
              * 33206 33966: contig of 761 bp in length
              * 33967 34066: gap of 100 bp
              * 34067 34816: contig of 750 bp in length
              * 34817 34916: gap of 100 bp
              * 34917 35662: contig of 746 bp in length
              * 35663 35762: gap of 100 bp
              * 35763 36534: contig of 772 bp in length
              * 36535 36634: gap of 100 bp
              * 36635 37399: contig of 765 bp in length
              * 37400 37499: gap of 100 bp
              * 37500 38246: contig of 747 bp in length
              * 38247 38346: gap of 100 bp
              * 38347 39098: contig of 752 bp in length

```



```

* 39099 39198: gap of 100 bp
* 39199 39937: contig of 739 bp in length
* 39938 40037: gap of 100 bp
* 40038 40783: contig of 746 bp in length
* 40784 40883: gap of 100 bp
* 40884 41572: contig of 789 bp in length
* 41573 41772: gap of 100 bp
* 41773 42535: contig of 763 bp in length
* 42536 42635: gap of 100 bp
* 42636 43403: contig of 768 bp in length
* 43404 43503: gap of 100 bp
* 43504 44255: contig of 752 bp in length
* 44256 44355: gap of 100 bp
* 44356 45086: contig of 731 bp in length
* 45087 45186: gap of 100 bp
* 45187 45916: contig of 730 bp in length
* 45917 46016: gap of 100 bp
* 46017 46741: contig of 725 bp in length
* 46742 46841: gap of 100 bp
* 46842 47597: contig of 756 bp in length
* 47598 47698: gap of 100 bp
* 47699 48450: contig of 753 bp in length
* 48451 48550: gap of 100 bp
* 48551 49303: contig of 753 bp in length
* 49304 49403: gap of 100 bp
* 49404 50159: contig of 756 bp in length
* 50160 50259: gap of 100 bp
* 50260 51011: contig of 752 bp in length
* 51012 51111: gap of 100 bp
* 51112 51600: contig of 749 bp in length
* 51601 51960: gap of 100 bp
* 51961 52723: contig of 763 bp in length
* 52724 52823: gap of 100 bp
* 52824 53589: contig of 766 bp in length
* 53590 53689: gap of 100 bp
* 53690 54439: contig of 750 bp in length
* 54440 54539: gap of 100 bp
* 54540 55302: contig of 763 bp in length
* 55303 55402: gap of 100 bp
* 55403 56157: contig of 755 bp in length
* 56158 56257: gap of 100 bp
* 56258 57008: contig of 751 bp in length
* 57009 57109: gap of 100 bp
* 57109 57880: contig of 772 bp in length
* 57881 57980: gap of 100 bp
* 57981 58742: contig of 762 bp in length
* 58743 58842: gap of 100 bp

```

```

Query Match      64.7% Score 19.4; DB 2; Length 68974;
Best Local Similarity 79.3%; Pred. No. 5.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Cy 1 CCTGCGTAGTGTGCTACGACCTCTGCGAG 29
Db 30913 CTCTGAGGTGTGCGAGACCTCTGCGAG 30885

RESULT 7
AC002315/c 110079 bp DNA linear ROD 20-FEB-1998
LOCUS Mouse BAC-146N21 Chromosome X contains iduronate-2-sulfatase gene;
DEFINITION complete sequence.
ACCESSION AC002315
VERSION AC002315.1 GI:2258164
KEYWORDS
SOURCE HTG.
ORGANISM Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 110079)
Muzny, D., Ansari-Lari, M.A., Timms, K.M., Yu, W., Dugan, S., Lu, J.,
Shen, Y., Kowland, K., Liu, W., Perez, L., Ding, Y., Haywood, M.,
Jain, A., Leal, B., Logan, O., Nguyen, V., Savage, L., Shen, H.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
COMMENT

```

Worley, K., Chen, E., Forcum, J., Atenson, A.D., Chiu, M.W., Gorrell, J.H., Brundage, E., Di, W., Chinnault, C., Nelson, D., and Gibbs, R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 110079)  
 Timms, K.M., Hackett, L., Edwards, F.J., Lu, J., Muzny, D.M., Miller, W., and Gibbs, R.A.  
 Cross-species sequence comparison of the IDS region  
 Unpublished  
 3 (bases 1 to 110079)  
 Chiu, M.W.  
 Direct Submission  
 Submitted (15-JUN-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

## FEATURES

## source

```

1..110079
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="BAC-146N21"
244..561
/rpc_family="LTR/MaLR"
repeat_region
567..2305
/rpc_family="LTR/MaLR"
repeat_region
2306..2646
/rpc_family="LTR/MaLR"
repeat_region
2647..3101
/rpc_family="LTR/Retroviral"
repeat_region
3102..3847
/rpc_family="Simple_repeat"
repeat_region
3848..5064
/rpc_family="Simple_repeat"
repeat_region
5065..5184
/rpc_family="SINE/B4"
repeat_region
5185..5457
/rpc_family="SINE/B2"
repeat_region
5458..6953
/rpc_family="SINE/B2"
repeat_region
6954..8199
/rpc_family="Simple_repeat"
repeat_region
8200..10633
/rpc_family="SINE/Alu"
repeat_region
10634..11185
/rpc_family="LTR/MaLR"
repeat_region
11186..11366
/rpc_family="SINE/B4"
repeat_region
11367..11718
/rpc_family="LTR/MaLR"
repeat_region
11719..11971
/rpc_family="Simple_repeat"
repeat_region
11972..12171
/rpc_family="SINE/B4"
repeat_region
12172..12396
/rpc_family="Low_complexity"
repeat_region
12397..12766
/rpc_family="SINE/B2"

```

```

repeat_region      14061..14118
                    /rpt_family="Simple_repeat"
unsnure             14096..14220
                    /note="single strand coverage only"
repeat_region      14179..14201
                    /rpt_family="Low_complexity"
gene                join(14201..14287,15145..15281,15929..16106,17952..18040,
                    18934..19134,20646..20816,25808..25934,27982..28155)
                    /gene="TDS"
repeat_region      complement(14424..14541)
                    /rpt_family="Simple_repeat"
repeat_region      complement(14783..14914)
                    /rpt_family="Simple_repeat"
repeat_region      17009..17084
                    /rpt_family="SINE/MIR"
repeat_region      complement(17555..17697)
                    /rpt_family="SINE/Alu"
repeat_region      complement(17620..17719)
                    /rpt_family="LTR/MaLR"
repeat_region      complement(17821..17869)
                    /rpt_family="SINE/MIR"
repeat_region      complement(18167..18314)
                    /rpt_family="SINE/Alu"
repeat_region      18371..18499
                    /rpt_family="SINE/B4"
repeat_region      19306..19411
                    /rpt_family="SINE/Alu"
repeat_region      complement(19902..20388)
                    /rpt_family="Other/MER21_gro"
repeat_region      complement(21582..21661)
                    /rpt_family="Simple_repeat"
repeat_region      complement(21662..21761)
                    /rpt_family="sacRNA"
repeat_region      complement(21684..21761)
                    /rpt_family="SINE/ID"
repeat_region      21812..21849
                    /rpt_family="DNA/MER1_type"
repeat_region      complement(23849..23940)
                    /rpt_family="SINE/Alu"
repeat_region      25152..25220
                    /rpt_family="LTR/MaLR"
repeat_region      complement(26165..26376)
                    /rpt_family="SINE/B2"
repeat_region      complement(26379..26577)
                    /rpt_family="SINE/B2"
repeat_region      26995..27125
                    /rpt_family="SINE/B2"
repeat_region      27322..27505
                    /rpt_family="SINE/B2"
repeat_region      complement(27695..27734)
                    /rpt_family="Low_complexity"
repeat_region      29292..30131
                    /rpt_family="LINE/L1"
repeat_region      30204..30233
                    /rpt_family="Simple_repeat"
repeat_region      complement(31179..31235)
                    /rpt_family="Simple_repeat"
repeat_region      31236..31266
                    /rpt_family="Simple_repeat"
repeat_region      31579..31703
                    /rpt_family="Simple_repeat"
repeat_region      31705..32083
                    /rpt_family="LINE/L1"
repeat_region      33920..33985
                    /rpt_family="Low_complexity"
repeat_region      34396..34534
                    /rpt_family="Simple_repeat"
repeat_region      complement(34591..34664)
                    /rpt_family="DNA/MER1_type"
repeat_region      complement(34824..34916)
                    /rpt_family="Simple_repeat"
repeat_region      complement(35493..35633)
                    /rpt_family="LINE/L1"

```

```

repeat_region      complement(36159..36380)
                    /rpt_family="SINE/B2"
repeat_region      complement(36381..36489)
                    /rpt_family="LTR/MaLR"
repeat_region      36490..36804
                    /rpt_family="LTR/MaLR"
repeat_region      36805..38233
                    /rpt_family="LTR/MaLR"
repeat_region      38236..38585
                    /rpt_family="LTR/MaLR"
repeat_region      complement(38986..39048)
                    /rpt_family="SINE/Alu"
repeat_region      39226..39295
                    /rpt_family="Low_complexity"
repeat_region      complement(39309..39438)
                    /rpt_family="SINE/Alu"
repeat_region      complement(39439..39929)
                    /rpt_family="LINE/L1"
repeat_region      complement(40689..40779)
                    /rpt_family="SINE/Alu"
repeat_region      complement(41671..42216)
                    /rpt_family="LTR/MaLR"
repeat_region      complement(42325..42428)
                    /rpt_family="Simple_repeat"
repeat_region      42539..42710
                    /rpt_family="Simple_repeat"
repeat_region      42898..42980
                    /rpt_family="Simple_repeat"
repeat_region      complement(43056..43106)
                    /rpt_family="Low_complexity"
repeat_region      43286..43340
                    /rpt_family="Low_complexity"
repeat_region      43406..43513
                    /rpt_family="SINE/Alu"
repeat_region      43947..44060
                    /rpt_family="Simple_repeat"
repeat_region      46601..46704
                    /rpt_family="DNA/MER1_type"
repeat_region      46977..47046
                    /rpt_family="Simple_repeat"

Query Match      64.7%; Score 19.4; DB 10; Length 110079;
Best Local Similarity 79.3%; Pred. No. 5.9e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  CCTGCGTAGTGTGACCTCGTCGAG 29
Db      4399  CACTGCTAGTGTGACCTCGTCGAG 4371

RESULT 8
LOCUS      AC121312
DEFINITION Mus musculus clone RP24-34305, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION      AC121312
VERSION      AC121312.3 GI:28975922
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 163132)
AUTHORS      Birren,B., Nuebaum,C. and Lander,E.
TITL      Mus musculus, clone RP24-34305
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 163132)
AUTHORS      Birren,B., Linton,L., Nuebaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhagater,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Deairelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,

```

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Kage, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kama, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R., Landers, T., Macdonald, P., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strassus, N., Sudramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 163132)

Birren, B., Nusbaum, C., Lander, E., Aboueleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., DeArillano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kama, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Minova, T., Mlenga, V., Murphy, J., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retter, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Stubbbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 17, 2003 this sequence version replaced gi:25956378.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L24754

Center clone name: 343\_J\_5

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 160194 bases at least Q40

Consensus quality: 161479 bases at least Q30

Consensus quality: 161963 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 162232; sum-of-contigs

Quality coverage: 8.4 in Q20 bases; agarose-fp

Quality coverage: 8.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as \* runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

\* as soon as it is available and the accession number will be preserved.

1 872: contig of 872 bp in length

873 972: gap of 100 bp

973 2003: contig of 1031 bp in length

2004 2103: gap of 100 bp

2104 4258: contig of 2155 bp in length

4259 4359: gap of 100 bp

4359 6590: contig of 2232 bp in length

6591 6690: gap of 100 bp

6691 14225: contig of 7535 bp in length

14226 14325: gap of 100 bp

14326 28588: contig of 14263 bp in length

28589 28689: gap of 100 bp

28690 44413: contig of 15725 bp in length

44414 44513: gap of 100 bp

44514 82435: contig of 37922 bp in length

82436 82535: gap of 100 bp

82536 127168: contig of 44633 bp in length

127169 127269: gap of 100 bp

127269 163132: contig of 35864 bp in length.

Location/Qualifiers

1.163132

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP24-3435"

/clone\_1b="RPCI-24 Male Mouse BAC"

1.872

/note="assembly\_fragment"

973.2003

/note="assembly\_fragment"

2104.4258

/note="assembly\_fragment"

4359.6590

/note="assembly\_fragment"

6591.14225

/note="assembly\_fragment"

14326.28588

/note="assembly\_fragment"

28689.44413

/note="assembly\_fragment"

44514.82435

/note="assembly\_fragment"

82536.127168

/note="assembly\_fragment"

127269.163132

/note="assembly\_fragment"

/note="assembly\_fragment"

clone\_end:79

vector\_side:right"

BASE COUNT 50534 a 33420 c 33393 g 44882 t 903 others

ORIGIN

Query Match 64.7%; Score 19.4; DB 2; Length 163132;

Best Local Similarity 79.3%; Pred. No. 6.1e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

2 CCGCGTAGTGTGACGACCTCTCGACGG 30

135188 CCGGCCAGGGTAGCACTTCGACGG 135216

RESULT 9

AC117138/c

AC117138

LOCUS

DEFINITION

unordered pieces.

ACCESSION

AC117138.5 GI:25073145

KEYWORDS

HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 176222)

Murphy, D., Maric, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D., Anyalbech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, K., Barnstead, M., Benahmed, F., Bielawski, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Eacotto, M., Eugene, C., Evans, C. A., Faller, T., Fan, G., Fernandez, S., Finley, M., Flaeg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howell, S., Hulyk, S., Hume, J., Idledit, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovacs, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenz, H., Louie, L., Louie, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, J., McLeod, M. P., McNally, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Neir, L., Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunola, G., Olariunsa, A., Pal, S., Parks, K., Patel, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldinger, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, C., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C., D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Swatek, A., Tabor, P., Taylor, C., Taylor, R., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valae, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wood, H., Worley, K., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 176222)

Worley, K. C.

Direct Submission

Submitted (08-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 176222)

Direct Submission

Rat Genome Sequencing Consortium.

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 19, 2002 this sequence version replaced gi:23813077. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

#### Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GUCH

Center clone name: CH230-385L3

Summary Statistics

Assembly program: Phrap, version 0.990329

Consensus quality: 163801 bases at least Q40

Consensus quality: 166235 bases at least Q30

Consensus quality: 167951 bases at least Q20

Estimated insert size: 166810; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).

NOTE: This is a 'working draft' sequence. It currently

consists of 4 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 3216: contig of 3216 bp in length

\* 3317 14915: gap of unknown length

\* 14916 15015: gap of unknown length

\* 15016 139403: contig of 124388 bp in length

\* 139404 139503: gap of unknown length

\* 139504 176222: contig of 36719 bp in length.

#### FEATURES

##### source

1. 176222

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-385L3"

1. 1160

/note="wgs\_end\_extension"

clone\_end:T7

3317. 4401

/note="wgs\_end\_extension"

clone\_end:T7

10164. 11253

/note="wgs\_end\_extension"

clone\_end:T7

13591. 14915

/note="wgs\_end\_extension"

clone\_end:T7

15016. 16802

/note="wgs\_end\_extension"

clone\_end:T7

17106. 18170

/note="wgs\_end\_extension"

clone\_end:T7

33714. 34597

/note="clone\_boundary"

clone\_end:T7

site:

end\_sequence:82134038"

BASE COUNT 50120 a 34724 c 34881 g 49698 t 6799 others

ORIGIN

Query Match 64.7% Score 19.4; DB 2; Length 176222;

Best Local Similarity 79.3%; Pred. No. 6.1e+02;

[illegible]





ACCESSION AC113777  
 VERSION AC113777.5 GI:25072715  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULFILLTOP.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 222869)  
 Mueny, D. Marie, Metzger, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D., Anulescu, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaru, D., Bandaru, D., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryan, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cessari, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dedecich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., George, G., Gier, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovacs, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louisedge, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, B., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Narkervic, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokenkeme, O., Okwonna, G., Olarnunagson, A., Pal, S., Parks, K., Patel, A., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Reiser, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wel, X., White, F., Williams, G., Wilson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiser, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 222869)  
 AUTHOR Morley, K. C.  
 JOURNAL Direct Submission  
 Submitted (05-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 222869)  
 Rat Genome Sequencing Consortium.  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 COMMENT On Nov 19, 2002 this sequence version replaced gi:23269985.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
 Center project name: GSGF  
 Center clone name: CH230-92115

----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 217955 bases at least Q40  
 Consensus quality: 218938 bases at least Q30  
 Consensus quality: 219754 bases at least Q20  
 Estimated insert size: 223449; sum-of-contigs estimation  
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 3 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

FEATURES  
 source  
 1. 222869  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-92115"  
 /note="wgs contig"  
 misc\_feature  
 59606 a 48726 c 48850 g 63215 t 2472 others  
 BASE COUNT  
 ORIGIN

Query Match 64.7%; Score 19.4; DB 2; Length 222869;  
 Best Local Similarity 79.3%; Pred. No. 6.2e+02;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 0y 1 CCTGCTAGTGTACGACCTCTGCAGG 29  
 Db 47915 CTCTGAGTTCTGGAGGACCTCTGCAGG 47943

RESULT 14  
 EX294168/c 226382 bp DNA linear HTG 02-JUN-2003  
 LOCUS Mus musculus chromosome X clone RP23-29M4, \*\*\* SEQUENCING IN  
 DEFINITION PROGRESS \*\*\*, 5 unordered pieces.  
 ACCESSION BX294168  
 VERSION BX294168.3 GI:11335441  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULFILLTOP.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:02:34 ; Search time 150.297 Seconds  
(without alignments)  
538.821 Million cell updates/sec

Title: US-09-594-065-3  
Perfect score: 30  
Sequence: 1 cccctgcgtacgtgctacgacctctccagagg 30

Scoring table: IDENTITY NUC  
Gapop 10-0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*

- 1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*
- 25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	62.7	576	22	AAH67112
C 2	18.8	62.7	1368	18	AAH67112
C 3	18.8	62.7	2936	18	AAH67112
C 4	18.8	62.7	4454	18	AAH67112
C 5	18.8	62.7	349980	22	AAH67112
C 6	18.4	61.3	294	24	ABN17942
C 7	18.4	61.3	3653	24	ABN19918
C 8	18.4	61.3	4705	24	ABA96146

9	18.4	61.3	4786	24	ABA96147
C 10	18.4	61.3	1503841	24	ABR001010
C 11	18.4	61.3	1503841	24	ABR01503
C 12	18.4	61.3	1503900	22	AAK95240
C 13	18.4	61.3	1503900	22	AAK95240
C 14	17.8	59.3	288	25	ABX82153
C 15	17.8	59.3	289	24	ABL72764
C 16	17.8	59.3	1035	22	AAH67112
C 17	17.8	59.3	1766	21	AAH67112
C 18	17.8	59.3	2035	15	AAH67112
C 19	17.8	59.3	2035	24	ABE62374
C 20	17.8	59.3	3444	23	ABL29461
C 21	17.8	59.3	9205	23	ABL29461
C 22	17.6	58.7	2553	23	ABL10891
C 23	17.6	58.7	4829	25	ABX08797
C 24	17.6	58.7	11431	23	ABL19365
C 25	17.6	58.7	21837	22	AAH67112
C 26	17.4	58.0	243	23	AAH67112
C 27	17.4	58.0	262	22	AAH67112
C 28	17.4	58.0	262	22	AAH67112
C 29	17.4	58.0	262	22	AAH67112
C 30	17.4	58.0	262	22	AAH67112
C 31	17.4	58.0	325	21	AAH67112
C 32	17.4	58.0	799	22	AAH67112
C 33	17.4	58.0	799	24	ABO6773
C 34	17.4	58.0	823	22	AAH08485
C 35	17.4	58.0	823	25	ABT22209
C 36	17.4	58.0	878	24	ABH53342
C 37	17.4	58.0	1062	24	ABL80201
C 38	17.4	58.0	1166	23	AAH67112
C 39	17.4	58.0	1446	21	AAH67112
C 40	17.4	58.0	1478	22	AAH67112
C 41	17.4	58.0	1487	22	AAH67112
C 42	17.4	58.0	1508	22	AAH67112
C 43	17.4	58.0	1520	22	AAH67112
C 44	17.4	58.0	1569	24	AAH67112
C 45	17.4	58.0	1592	23	AAH67112

## ALIGNMENTS

AAH67112/c	AAH67112 standard; DNA; 576 BP.
AC	AAH67112;
XX	26-SEP-2001 (first entry)
XX	C glutamicum coding sequence fragment SEQ ID NO: 2147.
XX	Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX	organic acid synthesis; de.
XX	Corynebacterium glutamicum.
XX	EP1108790-A2.
XX	20-JUN-2001.
XX	18-DEC-2000; 2000EP-0127688.
XX	16-DEC-1999; 99UP-0377484.
XX	07-APR-2000; 2000UP-0159162.
XX	03-AUG-2000; 2000UP-0280988.
XX	(KYOW ) KYOWA HAKKO KOGYO KK.
XX	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	WPI; 2001-376931/40.

Modified human/mou  
Human neurogulin 1  
Human neurogulin 1  
Human neurogulin-1  
Human neurogulin-1  
Corn ear-derived p  
Corn tassal-deriva  
P. putida oxygeas  
Human secreted pro  
Sequence encoding  
Colon adenocarcino  
Drosophila melanog  
Drosophila melanog  
Drosophila melanog  
Angiogenesis-assoc  
Drosophila melanog  
Human immune/haema  
DNA encoding novel  
Human immune/haema  
Human immune/haema  
Human immune/haema  
Human immune/haema  
Human immune/haema  
Human secreted pro  
Human cDNA encodin  
Human polynucleoti  
Human cDNA clone ( B  
Breast cancer mark  
Corn event MON180  
Human polynucleoti  
DNA encoding novel  
Eosinophil activat  
Human cytoskeletal  
Human polynucleoti  
Human polynucleoti  
Human polynucleoti  
Polynucleotide enc  
DNA encoding novel

DR P-P8DB; AAG91893.  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analyzing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 PS Claim 8; SEQ ID NO: 2147; 246bp + Sequence Listing; English.  
 CC  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of Coryneform bacterium, measuring expression amount and  
 CC analyzing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homolog of a gene derived  
 CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 CC  
 XX  
 SQ Sequence 576 BP, 101 A; 158 C; 148 G; 169 T; 0 other;  
 Query Match 62.7%; Score 18.8; DB 22; Length 576;  
 Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 9 AGTGTACGACCTCTGACGGG 30  
 DB 258 AGTGTACGACCACTGACAGG 237  
 ID AAT66463 standard; DNA; 1368 BP.  
 XX AAT66463;  
 AC  
 XX 25-MAR-2003 (updated)  
 DT 29-JUL-1997 (first entry)  
 XX  
 DE Thermophilic alkaline phosphatase gene.  
 XX  
 KM Alkaline phosphatase; thermophilic enzyme; label; assay; ss.  
 XX  
 OS Rhodothermus marinus strain ATCC 43812.  
 XX  
 OS Rhodothermus marinus strain ATCC 43812.  
 XX  
 FH Key Location/Qualifiers  
 FT 1..63  
 FT sig\_peptide /tag= a  
 FT /note= "the signal peptide is encoded by bases 1-60  
 FT mat\_peptide /tag= b  
 FT /note= "when the AP gene is expressed in E. coli"  
 FT 61..1365  
 FT /note= "the mature protein is encoded by bases  
 FT 61-1365 when the AP is expressed in E.  
 FT coli"  
 FT  
 PN EP770678-A2.  
 PD 02-MAY-1997.  
 XX  
 XX 24-OCT-1996; 96EP-0307692.  
 XX  
 XX 27-OCT-1995; 95US-0005965.  
 XX  
 PA (AMSH ) AMERSHAM LIFE SCI INC.  
 PA (AMSH ) AMERSHAM PHARMACIA BIOTECH INC.  
 XX  
 PI Davis M, Szaez J;  
 YX

DR WPI; 1997-238139/22.  
 DR P-P8DB; AAM17830.  
 XX  
 XX Thermotable Rhodothermus marinus alkaline phosphatase - useful as  
 PT enzyme label in immunoassays and nucleic acid assays  
 PT  
 XX  
 PS Example; Fig 9; 24pp; English.  
 CC  
 CC A genomic DNA sequence (AAT66461) from Rhodothermus marinus (Rma)  
 CC codes for a thermophilic alkaline phosphatase (AP) (AAM17830)  
 CC that shows optimal activity at pH 10.8 and which is stable to  
 CC heating at 65 deg for 1 hr. The sequence was deduced from a  
 CC clone isolated following expression screening of E. coli cells  
 CC transformed with Rma genomic DNA (see also AAT66461). The Rma  
 CC AP gene can be used to produce recombinant AP in transformed host  
 CC cells for use in nucleic acid and protein detection assays.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 XX  
 SQ Sequence 1368 BP, 218 A; 451 C; 486 G; 213 T; 0 other;  
 Query Match 62.7%; Score 18.8; DB 18; Length 1368;  
 Best Local Similarity 76.7%; Pred. No. 1.4e+02;  
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 CCTGCGTAGTGACGACCTCTGACGGG 30  
 DB 197 CGTCCGTACTGCTTCCGCTCCAGCCGGG 226  
 ID AAT66461 standard; DNA; 2936 BP.  
 XX AAT66461;  
 AC  
 XX 25-MAR-2003 (updated)  
 DT 29-JUL-1997 (first entry)  
 XX  
 DE Thermophilic alkaline phosphatase gene.  
 XX  
 KM Alkaline phosphatase; thermophilic enzyme; label; assay; ss.  
 XX  
 OS Rhodothermus marinus strain ATCC 43812.  
 XX  
 OS Rhodothermus marinus strain ATCC 43812.  
 XX  
 PN EP770678-A2.  
 PD 02-MAY-1997.  
 XX  
 XX 24-OCT-1996; 96EP-0307692.  
 XX  
 XX 27-OCT-1995; 95US-0005965.  
 XX  
 PA (AMSH ) AMERSHAM LIFE SCI INC.  
 PA (AMSH ) AMERSHAM PHARMACIA BIOTECH INC.  
 XX  
 PI Davis M, Szaez J;  
 WPI; 1997-238139/22.  
 XX  
 PT Thermotable Rhodothermus marinus alkaline phosphatase - useful as  
 PT enzyme label in immunoassays and nucleic acid assays  
 XX  
 PS Example; Fig 7; 24pp; English.  
 CC  
 CC A genomic DNA sequence (AAT66461) from Rhodothermus marinus (Rma)  
 CC contains the gene for a thermophilic alkaline phosphatase (AP) and  
 CC flanking regions. It was obtained by transforming E. coli XL2-Blue  
 CC MRF<sup>+</sup> cells with Rma genomic DNA fragments, screening for AP  
 CC activity on plates using a chromogenic indicator. and KpnI  
 CC digestion of isolated clone pBRM1. A 1.6 kb fragment of pBRM1/KpnI  
 CC was subcloned into pUC19 to yield plasmid pBRM1.8 (AAT66462), and  
 CC a presumptive full-length coding sequence (AAT66463) for Rma AP  
 CC (AAM17830) was produced. The Rma AP nucleic acids can be used for

CC prodn. of recombinant AP for use in nucleic acid and protein  
CC detection assays.  
CC (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 2936 BP; 525 A; 1012 C; 880 G; 519 T; 0 other;

Query Match 62.7%; Score 18.8; DB 18; Length 2936;  
Best Local Similarity 76.7%; Pred. No. 1.5e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCCGCGTAGTCGTACGACCTCTGCAGG 30  
Db 1351 CCGTGGTACTGCTTCGCGCTTCAGCGGG 1322

## RESULT 4

AAT66462/c  
ID AAT66462 standard; DNA; 4454 BP.

XX AAT66462;

XX 25-MAR-2003 (updated)  
DT 25-JUN-1997 (first entry)

XX Plasmid pCRM1.8 containing an alkaline phosphatase gene.

XX Alkaline phosphatase; thermophilic enzyme; label; assay;  
KW plasmid pCRM1.8; ss.

XX Chimeric Rhodothermus marinus strain ATCC 43812.

XX EP70678-A2.

XX 02-MAY-1997.

XX 24-OCT-1996; 96EP-0307692.

XX 27-OCT-1995; 95US-0005965.

XX (AMSH ) AMERSHAM LIFE SCI INC.  
PA (AMSH ) AMERSHAM PHARMACIA BIOTECH INC.

XX Davis M, Szasz J;

XX WPI; 1997-238139/22.

XX Thermostable Rhodothermus marinus alkaline phosphatase - useful as  
PT enzyme label in immunoassays and nucleic acid assays

XX Example; Fig 8; 24pp; English.

XX Plasmid pCRM1.8 (AAT66463) contains a 1.8 kb HindIII/KpnI segment of  
CC genomic DNA (see also AAT66461) containing the alkaline phosphatase  
CC (AP) and flanking sequences from Rhodothermus marinus (Rma), cloned  
CC into pUC19. This recombinant plasmid has been used to produce Rma  
CC thermophilic AP in E. coli JM109 transformants. A presumptive  
CC full-length coding sequence (AAT66463) for Rma AP (AAW17830) was  
CC deduced. The Rma AP nucleic acids can be used for prodn. of  
CC recombinant AP for use in nucleic acid and protein detection  
CC assays.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 4454 BP; 978 A; 1305 C; 1234 G; 937 T; 0 other;

Query Match 62.7%; Score 18.8; DB 18; Length 4454;  
Best Local Similarity 76.7%; Pred. No. 1.5e+02;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCCGCGTAGTCGTACGACCTCTGCAGG 30  
Db 1738 CCGTGGTACTGCTTCGCGCTTCAGCGGG 1709

RESULT 5  
AAH68530/c  
ID AAH68530 standard; DNA; 34980 BP.

XX AAH68530;

XX 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 7065.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159152.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
XX Disclosure; SEQ ID NO: 7065; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of Coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from Coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 349980 BP; 86896 A; 98023 C; 80939 G; 84122 T; 0 other;

Query Match 62.7%; Score 18.8; DB 22; Length 349980;  
Best Local Similarity 90.9%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 9 AGGTGACGACCTCTGCAGG 30  
Db 266823 AGGTGACGACCTCTGCAGG 266802

RESULT 6  
ABN17942  
ID ABN17942 standard; cDNA; 294 BP.

XX ABN17942;

XX 24-JUN-2002 (first entry)

XX Human ORF polynucleotide sequence SEQ ID NO:4361.

KM		Human; open reading frame; ORFX; gene therapy; Cancer; cirrhosis;
KM		hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KM		degenerative disorder; osteoarthritis; neurodegenerative disorder;
KM		cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KM		hypertension; hypothyroidism; cholesterol ester storage disease;
KM		immune deficiency; immune disorder; infectious disease;
KM		autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KM		myasthenia gravis; gene; ss.
XX		
OS	Homo sapiens.	
PN	MOL00192523-A2.	
XX		
PD	06-DEC-2001.	
XX		
PF	29-MAY-2001; 2001MO-US10836.	
PR	30-MAY-2000; 2000US-206132P.	
FR	29-AUG-2000; 2000US-228716P.	
XX		
PA	(CURA-) CURAGEN CORP.	
PI	Shimkets RA, Leach MD;	
DR	WPI, 2002-106308/14.	
DR	p-PDSB; ABP02190.	
PT	Noval human polypeptides and polymucleotides useful for diagnosing,	
PT	preventing and treating cardiovascular disease, neurodegenerative,	
XX	hyperproliferative disorders and autoimmune disorders -	
XX		
PS	Disclosure; SEQ ID 4361; 1037pp; English.	
CC	The present invention describes substantially purified human proteins	
CC	(referred to as open reading frame, ORFX, where X is I-I1491 (see Table 1	
CC	in the specification)). AEN15762 to AEN27252 encode the human ORFX	
CC	protein given in ABP00010 to ABP11500. ORFX proteins are useful for	
CC	treating or preventing a pathology associated with an ORFX-associated	
CC	disorder in humans, and in the manufacture of a medicament for treating a	
CC	syndrome associated with ORFX-associated disorder. ORFX polymucotide	
CC	sequences can be used in gene therapy. ORFX sequences can be used in the	
CC	treatment of cancer, hyperproliferative disorders, cirrhosis of liver,	
CC	psoriasis, benign tumours, keloids, degenerative disorders, haemorrhage,	
CC	osteoarthritis, neurodegenerative disorders, disorders related to organ	
CC	transplantation, cardiovascular diseases, diabetes mellitus, systemic	
CC	lupus erythematosus, hypertension, hypothyroidism, cholesterol ester	
CC	storage disease, various immune deficiencies and disorders, infectious	
CC	diseases, autoimmune disorders such as multiple sclerosis, rheumatoid	
CC	arthritis, autoimmune thyroditis, myasthenia gravis, graft-versus-host	
CC	disease and autoimmune inflammatory eye disease. ORFX proteins are also	
CC	useful for treating burns, incisions, ulcers, for treating osteoporosis,	
CC	bone degenerative disorders, or periodontal disease, and for gut	
CC	protection or regeneration and treatment of lung or liver fibrosis,	
CC	reflexion injury in various tissues and conditions resulting from	
CC	systemic cytokine damage.	
CC	N.B. The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pot_sequences.	
XX		
SQ	Sequence 294 BP; 91 A; 63 C; 82 G; 56 T; 2 other;	
	Query Match            61.3%; Score 18.4; DB 24; Length 294;	
	Best Local Similarity 78.6%; Pred. No. 1.9e+02;	
	Matches     22; Conservative      0; Mismatches        6; Indels          0; Gaps            0	
Oy	CTGCGTAGTGGTAAGACATTCCTGCAGGG 30       	
Db	248 CTGGGTGTGGTTCAACATCCTCACAGG 275       	

RESULT\_7  
AB199918

ID	AB199918	standard; cDNA; 3653 BP.
XX		
AC	AB199918;	
XX		
DT	08-MAR-2002	(first entry)
XX		
DE	Rat mucocardial cell proliferation associated cDNA seq ID NO 7.	
XX		
KM	Rat; heart; cardiac; myocardial necrosis; cardiac hypertrophy;	
KW	cardiac insufficiency; ss.	
XX		
OS	Rattus norvegicus.	
XX		
PN	NO200183705-A1.	
XX		
PD	08-NOV-2001.	
XX		
PF	27-APR-2001; 2001WO-JP03700.	
XX		
PR	27-APR-2000; 2000JP-0126741.	
XX		
PA	(KYOW ) KYOWA HAKKO KOGYO KK.	
XX		
PI	Yamada Y, Sekine S, Kikuchi Y, Sakurada K;	
XX		
DR	WPI; 2002-075160/10.	
XX		
PT	P-PSDB; ABB57378.	
XX		
PT	Genes having differential expression in fetal and adult heart tissue	
XX	useful for screening potential drugs for promoting repair of damage	
XX	caused by myocardial necrosis	
PS	Claim 22; Page 84-90; 171pp; Japanese.	
XX		
CC	The invention relates to gene sequences (AB199915-AB199934) having	
CC	modified expression in fetal heart tissue as compared to adult heart	
CC	tissue and the encoded proteins (ABB57375-ABB57397). The genes have	
CC	cardiant activity and may be useful in the promotion of the repair of	
CC	damage to heart tissue caused by myocardial necrosis. The gene sequences	
CC	are useful for screening potential compounds for the ability to influence	
CC	disease associated with myocardial necrosis. Drugs identified by the	
CC	screening methods may be used to treat and prevent disease with which	
CC	myocardial necrosis is associated, such as cardiac hypertrophy and	
CC	cardiac insufficiency. Diagnosis of diseases such as those above is also	
CC	disclosed.	
XX		
SO	Sequence 3653 BP; 845 A; 1090 C; 1048 G; 670 T; 0 other;	
	Query Match	61.3%; Score 18.4; DB 24; Length 3653;
	Best Local Similarity	78.6%; Pred. No. 2,3e+02;
	Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0	
OY	2 CCTGCGTAGTGCAGCCTCTGCAGG 29	
DB	1788 CTGCGTTCTGTGGCCTCTGCATG 1815	
RESULT 8		
ABA96146		
ID	ABA96146	standard; DNA; 4705 BP.
XX		
AC	ABA96146;	
XX		
DT	15-APR-2002	(first entry)
XX		
DE	Modified human/mouse Bcr-Abi fusion gene #1.	
XX		
KW	Human; mouse; Bcr-Abi; cancer; cytostatic; immunosuppressive; leukaemia;	
KW	orthologous bone marrow transplantation; ds.	
XX		
OS	Chimeric - Homo sapiens.	
XX		
OS	Chimeric - Mus musculus.	
XX		

FH	Key	Location/Qualifiers
FT	CDS	/tag= 1..4704
FT		a
FT		/product= "Bcr-Abl fusion protein #1"
XX		
FN	WO20020024-A1.	
XX		
PD	03-JAN-2002.	
XX		
PF	29-JUN-2001; 2001WO-US20602.	
XX		
PR	30-JUN-2000; 2000US-215595P.	
XX		
PA	(REGC ) UNIV CALIFORNIA.	
XX		
P1	Wang JYJ, Vigneri P;	
DR	MPI; 2002-139846/18.	
DR	P-PSDB; ABB08236.	
XX		
PT	Killing cancer cells for treating leukaemia, involves causing accumulation of activated Bcr-Abl in the nucleus to induce apoptosis -	
PS	Disclosure; Page 33-40; 63pp; English.	
XX		
CC	The sequence encodes a human/mouse Bcr-Abl fusion protein. The invention relates to a novel method for killing cancer cells, comprising causing an accumulation of activated Bcr-Abl in the nucleus, to induce apoptosis. The method has cytostatic and immunosuppressive activity. The method is useful for killing cancer cells, and in the treatment of cancers including leukaemia. The invention also supplies a method useful for purging bone marrow to allow for orthologous bone marrow transplantation, and as a therapeutic option for chronic myelogenous leukaemia.	
XX		
SQ	Sequence 4705 BP; 1072 A; 1457 C; 1435 G; 741 T; 0 other;	
QY	Query Match 61.3%; Score 18.4; DB 24; Length 4705; Best Local Similarity 78.6%; Pred.No. 2.3e+02; Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0,	
Dh	2 CCTGGCTAGTGTGACGACTCTCTGCAGC 29                           3414 CTGGCTTGTTGCGGCTCTCGCATG 3441	
RESULT 9		
ID	ABA96147 standard; DNA; 4786 BP.	
XX		
AC	ABA96147;	
XX		
DT	15-APR-2002 (first entry)	
XX		
DB	Modified human/mouse Bcr-Abl fusion gene #2.	
XX		
KM	Human; mouse; Bcr-Abl; cancer; cytostatic; immunosuppressive; leukemia;	
OS	orthologous bone marrow transplantation; ds.	
XX		
OS	Chimeric - Homo sapiens.	
XX		
OS	Chimeric - Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	/tag= 1..4785
FT		a
FT		/product= "Bcr-Abl fusion protein #2"
XX		
XX		
PN	WO20020024-A1.	
XX		
PD	03-JAN-2002.	
XX		
PF	29-JUN-2001; 2001WO-US20602.	
XX		

PR	30-JUN-2000; 2000US-215595P.
PA	(REGC ) UNIV CALIFORNIA.
PB	
PI	Wang JYJ, Vigneri P;
PL	
PM	MPJ; 2002-139846/18.
PN	P-PSTD; ABB08237.
PO	
PP	Killing cancer cells for treating leukaemia, involves causing accumulation of activated Bcr-Abl in the nucleus to induce apoptosis -
PQ	
PS	Example 10; Page 47-54; 63pp; English.
PT	
PV	
PX	The sequence encodes a human/mouse Bcr-Abl fusion protein. The invention relates to a novel method for killing cancer cells, comprising causing an accumulation of activated Bcr-Abl in the nucleus, to induce apoptosis. The method has cytostatic and immunosuppressive activity. The method is useful for killing cancer cells, and in the treatment of cancers including leukemia. The invention also supplies a method useful for purging bone marrow to allow for orthologous bone marrow transplantation, and as a therapeutic option for chronic myelogenous leukaemia.
QA	
QB	
QC	Sequence 4786 BP; 1104 A; 1470 C; 1465 G; 747 T; 0 other:
QD	
QE	Query Match 61.3%; Score 18.4; DB 24; Length 4786;
QF	Best Local Similarity 78.6%; Pred.No. 2.3e+02;
QG	Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QH	
QI	2 CCTGCGTAGTGGCTACGACCTCTTCGCAAG 29
QJ	
QK	Db 3414 CCTGCGTTCTGTTCGGCCTCTGCATG 3441
QL	
QM	
QN	
QO	
QP	
QQ	
QR	
QS	
QT	
QU	
QV	
QW	
QX	
QY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	
RG	
RH	
RI	
RJ	
RK	
RL	
RM	
RN	
RO	
RP	
RQ	
RR	
RS	
RT	
RU	
RV	
RW	
RX	
RY	
RA	
RB	
RC	
RD	
RE	
RF	

```

FT /number= 3
FT exon 826308..826355
FT /tag= h
FT /number= 4
FT intron 826356..1034242
FT /tag= i
FT /number= 4
FT exon 1034243..1034321
FT /tag= j
FT /number= 5
FT intron 1034322..1153294
FT /tag= k
FT /number= 5
FT exon 1153295..1153886
FT /tag= l
FT /number= 6
FT intron 1153887..1200721
FT /tag= m
FT /number= 6
FT exon 1200722..1201065
FT /tag= n
FT /number= 7a
FT intron 1153887..1200887
FT /tag= o
FT /number= 7a
FT exon 1200888..1201065
FT /tag= p
FT /number= 7b
FT intron 1201066..1210622
FT /tag= q
FT /number= 7b
FT exon 1210623..1210744
FT /tag= r
FT /number= 8
FT intron 1210745..1219542
FT /tag= s
FT /number= 8
FT exon 1219543..1219593
FT /tag= t
FT /number= 9
FT intron 1219594..1221863
FT /tag= u
FT /number= 9
FT exon 1221864..1221914
FT /tag= v
FT intron 1221915..1252253
FT /tag= w
FT /number= 10
FT exon 1252254..1253413
FT /tag= x
FT /number= 11
FT intron 1253414..1326781
FT /tag= y
FT /number= 11
FT exon 1326782..1337071
FT /tag= z
FT /number= 12
FT intron 1337072..1332977
FT /tag= aa
FT /number= 12
FT exon 1332978..1333107
FT /tag= ab
FT /number= 13a
FT intron 1333108..1347039
FT /tag= ac
FT /number= 13a
FT exon 1332978..1333652
FT /tag= ad
FT /number= 13b
FT intron 1333653..1347039
FT /tag= ae
FT /number= 13b

```

```

FT exon 1347040..1347107
FT /tag= af
FT /number= 14
FT intron 1347108..1347706
FT /tag= ag
FT /number= 14
FT exon 1347707..1347765
FT /tag= ah
FT /number= 15a
FT intron 1347766..1354620
FT /tag= aj
FT /number= 15a
FT exon 1347707..1348257
FT /tag= ak
FT /number= 15b
FT intron 1348258..1354620
FT /tag= al
FT /number= 15b
FT exon 1354621..1354644
FT /tag= am
FT /number= 16
FT intron 1354645..1359431
FT /tag= an
FT /number= 16
FT exon 1359432..1359534
FT /tag= ao
FT /number= 17
FT intron 1359535..1361462
FT /tag= ap
FT /number= 17
FT exon 1361463..1361589
FT /tag= aq
FT /number= 18
FT intron 1361590..1364362
FT /tag= ar
FT /number= 18
FT exon 1364363..1364493
FT /tag= as
FT /number= 19
FT intron 1364494..1365256
FT /tag= at
FT /number= 19
FT exon 1365257..1365463
FT /tag= au
FT /number= 20a
FT intron 1365464..1368280
FT /tag= av
FT /number= 20a
FT exon 1365257..1366044
FT /tag= aw
FT /number= 20b
FT intron 1366045..1368280
FT /tag= ax
FT /number= 20b
FT exon 1368281..1368422
FT /tag= ay
FT /number= 21
FT intron 1368423..1368810
FT /tag= az
FT /number= 21
FT exon 1368811..1369656
FT /tag= ba
FT /number= 22
FT CDS 629728..744437
FT /tag= bb
FT /product= "neuregulin-1-associated gene 1"
FT /note= "this sequence contains introns"
FT exon 629728..629804
FT /tag= bc
FT /number= 1
FT intron 629805..630974
FT /tag= bd
FT /number= 1

```



```
FT exon 630975..631187
FT /*tag= be
FT /number= 2
FT intron 631188..631282
FT /*tag= bf
FT /number= 2
FT exon 631283..631320
FT /*tag= bg
FT /number= 3
FT intron 631321..634334
FT /*tag= bh
FT exon 634335..634441
FT /*tag= bi
FT /number= 4a
FT intron 634442..635331
FT /*tag= bj
FT /number= 4a

Query Match 61.3%; Score 18.4; DB 24; Length 1503841;
Best Local Similarity 78.6%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Cy 1 CCTGCGTAGTGTGACGACCTCTGCAG 28
Db 801369 CCTGACTAGTGGAGACAGCTGCAG 801342

RESULT 11
ABT01503/c
ID ABT01503 standard; DNA; 1503841 BP.
XX AC ABT01503;
XX DT 07-NOV-2002 (first entry)
XX DE Human neuregulin 1 gene.
XX KW Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
XX KW schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
XX KW neuroleptic; gene therapy; gene; ds.
XX OS Homo sapiens.
XX Key Location/Qualifiers
FT CDS 244312..1369465
FT /*tag= a
FT /product= "neuregulin 1"
FT /note= "this sequence contains introns"
FT exon 244205..244348
FT /*tag= b
FT /number= 1
FT intron 244349..244640
FT /*tag= c
FT /number= 1
FT exon 244641..245646
FT /*tag= d
FT /number= 2
FT intron 245647..826009
FT /*tag= e
FT /number= 2
FT exon 826010..826101
FT /*tag= f
FT /number= 3
FT intron 826102..826307
FT /*tag= g
FT exon 826308..826355
FT /*tag= h
FT /number= 4
FT intron 826356..1034242
FT /*tag= i
FT /number= 4
```

```
FT exon 1034243..1034321
FT /*tag= j
FT /number= 5
FT intron 1034322..1153294
FT /*tag= k
FT /number= 5
FT exon 1153295..1153886
FT /*tag= l
FT /number= 6
FT intron 1153887..1200721
FT /*tag= m
FT /number= 6
FT exon 1200722..1201065
FT /*tag= n
FT /number= 7a
FT intron 1153887..1200887
FT /*tag= o
FT /number= 7a
FT exon 1200888..1201065
FT /*tag= p
FT /number= 7b
FT intron 1201066..1210622
FT /*tag= q
FT /number= 7b
FT exon 1210623..1210744
FT /*tag= r
FT /number= 8
FT intron 1210745..1219542
FT /*tag= s
FT /number= 8
FT exon 1219543..1219593
FT /*tag= t
FT /number= 9
FT intron 1219594..1221863
FT /*tag= u
FT /number= 9
FT exon 1221864..1221914
FT /*tag= v
FT /number= 10
FT intron 1221915..1252253
FT /*tag= w
FT /number= 10
FT exon 1252254..1253413
FT /*tag= x
FT /number= 11
FT intron 1253414..1326781
FT /*tag= y
FT /number= 11
FT exon 1326782..1327071
FT /*tag= z
FT /number= 12
FT intron 1327072..1332977
FT /*tag= aa
FT /number= 12
FT exon 1332978..1333107
FT /*tag= ab
FT /number= 13a
FT intron 1333108..1347039
FT /*tag= ac
FT /number= 13a
FT exon 1332978..1333652
FT /*tag= ad
FT /number= 13b
FT intron 1333653..1347039
FT /*tag= ae
FT /number= 13b
FT exon 1347040..1347107
FT /*tag= af
FT /number= 14
FT intron 1347108..1347706
FT /*tag= ag
FT /number= 14
FT exon 1347707..1347765
```

```

FT      /*tag= ah
FT      /number= 15a
FT      1347766..1354620
FT      /*tag= aj
FT      /number= 15a
FT      1347707..1348257
FT      /*tag= ak
FT      /number= 15b
FT      1348258..1354620
FT      /*tag= al
FT      /number= 15b
FT      1354621..1354644
FT      /*tag= am
FT      /number= 16
FT      1354645..1359431
FT      /*tag= an
FT      /number= 16
FT      1359432..1359534
FT      /*tag= ao
FT      /number= 17
FT      1359535..1361462
FT      /*tag= ap
FT      /number= 17
FT      1361463..1361589
FT      /*tag= aq
FT      /number= 18
FT      1361590..1364362
FT      /*tag= ar
FT      /number= 18
FT      1364363..1364493
FT      /*tag= as
FT      /number= 19
FT      1364494..1365256
FT      /*tag= at
FT      /number= 19
FT      1365257..1365463
FT      /*tag= au
FT      /number= 20a
FT      1365464..1368280
FT      /*tag= av
FT      /number= 20a
FT      1365257..1366044
FT      /*tag= aw
FT      /number= 20b
FT      1366045..1368280
FT      /*tag= ax
FT      /number= 20b
FT      1368281..1368422
FT      /*tag= ay
FT      /number= 21
FT      1368423..1368810
FT      /*tag= az
FT      /number= 21
FT      1368811..1369656
FT      /*tag= ba
FT      /number= 22
FT      629728..744437
FT      /*tag= bb
FT      /product= "neuregulin-1-associated gene 1"
FT      /note= "this sequence contains introns"
FT      629728..629804
FT      /*tag= bc
FT      /number= 1
FT      629805..630974
FT      /*tag= bd
FT      /number= 1
FT      630975..631187
FT      /*tag= be
FT      /number= 2
FT      631188..631282
FT      /*tag= bf
FT      /number= 2
FT      631283..631320
FT      exon

```

```

FT      /*tag= bg
FT      /number= 3
FT      631321..634334
FT      /*tag= bh
FT      /number= 3
FT      634335..634441
FT      /*tag= bi
FT      /number= 4a
FT      634442..635331
FT      /*tag= bj
FT      /number= 4a
FT      intron

Query Match      61.3%; Score 18.4; DB 24; Length 1503841;
Best Local Similarity 78.6%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY      1 CCTGCGTAGTGTGACGACCTCTGCAG 28
DB      801369 CCTGACTAGTGGAGACAGCCTGCAG 801342

RESULT 12
AAK95240/c
ID      AAK95240 standard; DNA; 1503900 BP.
XX
AC      AAK95240;
XX
DT      17-DEC-2001 (first entry)
XX
DE      Human neuregulin-1 gene.
XX
KW      Human; neuregulin-1 associated gene 1; NRGIAG1; schizophrenia gene;
XX      gene therapy; ds.
XX
OS      Homo sapiens.
XX
PN      WO200164876-A2.
XX
PD      07-SEP-2001.
XX
PF      28-FEB-2001; 2001WO-US06376.
XX
PR      28-FEB-2000; 2000US-0515715.
XX
PA      (DECO-) DECODE GENETICS EHF.
XX
PI      Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR      WPI; 2001-550179/61.
XX
DR      P-PGDB; AAG67900, AAG67901, AAG67902, AAG67903, AAG67904, AAG67905,
XX      AAG67906, AAG67907, AAG67908, AAG67909, AAG67910, AAG67911, AAG67912,
XX      AAG67913, AAG67914, AAG67915, AAG67916, AAG67917, AAG67918, AAG67919,
XX      AAG67920, AAG67921, AAG67922, AAG67923, AAG67924, AAG67925, AAG67926,
XX      AAG67927, AAG67928, AAG67929, AAG67930, AAG67931, AAG67932, AAG67933,
XX      AAG67934, AAG67935, AAG67936, AAG67937.
XX
PT      Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
XX      preventing diagnosing and creating schizophrenia -
XX
PS      Disclosure; Page 90-501; 750pp; English.
XX
CC      This sequence represents the human neuregulin-1 associated gene 1
XX      (NRGIAG1) of the invention. The NRGIAG1 gene is also referred to as the
XX      human Schizophrenia gene. The invention also relates to fragments or
XX      variants of the gene and the NRGIAG1 polypeptides they encode. The
XX      NRGIAG1 nucleic acids and polypeptides may be used in the prevention,
XX      diagnosis and treatment of diseases associated with inappropriate NRGIAG1
XX      expression. For example, they may be used to treat disorders associated
XX      with decreased expression by rectifying mutations or deletions in a
XX      patient's genome that affect the activity of NRGIAG1 by expressing
XX      inactive proteins or to supplement the patients own production of
XX      NRGIAG1. Additionally, the gene may be used to produce NRGIAG1
XX      polypeptides, by inserting the nucleic acids into a host cell and

```

CC culturing the cell to express the protein. The gene may also be used as  
CC DNA probes and primers in diagnostic assays to detect and quantitate the  
CC presence of similar nucleic acids in samples, and therefore which  
CC patients may be in need of restorative therapy. The NRG1AG1 polypeptides  
CC may also be used as antigens in the production of antibodies against  
CC NRG1AG1 and in assays to identify modulators of NRG1AG1 expression and  
CC activity. Anti-NRG1AG1 antibodies and antagonists may also be used to  
CC down regulate expression and activity. Anti-NRG1AG1 antibodies may  
CC also be used as diagnostic agents for detecting the presence of NRG1AG1  
CC polypeptides in samples. NRG1AG1 is associated with schizophrenia which  
CC may be prevented, diagnosed and/or treated by the above methods.

SO Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;

Query Match 61.3%; Score 18.4; DB 22; Length 1503900;  
Best Local Similarity 78.6%; Pred. No. 3.1e+02;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CCTGCGTAGTGTAACGACCTCTGCAG 28  
Db 801368 CCTGCGTAGTGTAACGACCTCTGCAG 801341

RESULT 13  
AAK96733/c  
ID AAK96733 standard; DNA; 1503900 BP.

XX AAK96733;

XX 17-DEC-2001 (first entry)

XX Human neuregulin-1 gene.

XX Human; neuregulin 1 gene; schizophrenia; gene therapy; de.

XX Homo sapiens.

XX MO200164877-A2.

XX 07-SEP-2001.

XX 28-FEB-2001; 2001WO-US06377.

XX 28-FEB-2000; 2000US-0515716.

XX (DECO-) DECODE GENETICS EHF.

XX Stefansson H, Steinthorsdottir V, Gulcher JR;

XX WPI; 2001-514841/56.

DR P-PSDB; AAG67938, AAG67939, AAG67940, AAG67941, AAG67942, AAG67943,  
DR AAG67944, AAG67945, AAG67946, AAG67947, AAG67948, AAG67949,  
DR AAG67950, AAG67951, AAG67952, AAG67953, AAG67954, AAG67955,  
DR AAG67956, AAG67957, AAG67958, AAG67959, AAG67960, AAG67961,  
DR AAG67962, AAG67963, AAG67964, AAG67965, AAG67966, AAG67967,  
DR AAG67968, AAG67969, AAG67970, AAG67971, AAG67972, AAG67973,  
DR AAG67974, AAG67975.

XX Neuregulin 1 nucleic acids and proteins useful for diagnosing

PT preventing and treating schizophrenia -

XX Disclosure; Page 345-756; 756pp; English.

CC This sequence represents the human neuregulin 1 gene of the invention.  
CC The invention also relates to fragments or variants of the neuregulin 1  
CC gene. The gene and its proteins may be used in the prevention, diagnosis  
CC and treatment of diseases associated with inappropriate neuregulin 1  
CC expression, such as schizophrenia. For example they may be used to treat  
CC disorders associated with decreased neuregulin 1 expression by rectifying  
CC mutations or deletions in a patient's genome that affect the activity of  
CC neuregulin 1 by expressing inactive proteins or to supplement the  
CC patients own production of polypeptides. Additionally, the gene may be  
CC used to produce the neuregulin 1 protein, by inserting the nucleic acids

CC into a host cell and culturing the cell to express the protein. The gene  
CC and its complementary sequences may also be used as DNA probes in  
CC diagnostic assays to detect and quantitate the presence of similar  
CC nucleic acids in samples, and therefore which patients may be in need of  
CC restorative therapy. The protein may also be used as antigens in the  
CC production of antibodies against neuregulin 1 and in assays to identify  
CC modulators of neuregulin 1 expression and activity. The antibodies and  
CC antagonists may also be used to down regulate expression and activity.  
CC The antibodies may also be used as diagnostic agents for detecting the  
CC presence of neuregulin 1 in samples.

SO Sequence 1503900 BP; 452487 A; 281874 C; 288074 G; 480092 T; 1373 other;

Query Match 61.3%; Score 18.4; DB 22; Length 1503900;  
Best Local Similarity 78.6%; Pred. No. 3.1e+02;  
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CCTGCGTAGTGTAACGACCTCTGCAG 28  
Db 801368 CCTGCGTAGTGTAACGACCTCTGCAG 801341

RESULT 14  
ABX82153

ID ABX82153 standard; cDNA; 288 BP.

XX ABX82153;

XX 24-APR-2003 (first entry)

XX Corn ear-derived polynucleotide (cpd) #613.

XX Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022;

XX SATMON023; structural gene; functional gene; regulatory gene;

XX corn ear-specific profile; gene transcription; gene expression;

XX hybrid plant; desirable trait expression; plant breeding program;

XX inheritance; desired characteristic; growth; development;

XX disease resistance; environmental adaptability; quality; yield;

XX multigene trait; plant; gene; ss.

XX Zea mays.

XX US6476212-B1.

XX 05-NOV-2002.

XX 14-MAY-1999; 99US-0313294.

XX 26-MAY-1998; 98US-086722P.

XX (INCY-) INCYTE GENOMICS INC.

XX Laijudi RV, Ito LY, Sherman BK;

XX WPI; 2003-208840/20.

XX Novel purified corn-ear derived polynucleotide useful as hybridization

PT probe for detecting polynucleotide in sample, and for identifying,

PT evaluating, and altering desired characteristics associated with

PT growth, development -

XX Examples; SEQ ID NO 613; 390pp; English.

CC The present invention relates to the isolation of corn ear-derived  
CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries  
CC SATMON022 and SATMON023. Some of the cpds uniquely identify structural,  
CC functional, and regulatory genes of corn ear. The polynucleotides  
CC sequences are useful for detecting cpds in a sample, for producing  
CC a corn ear-specific profile of gene transcription, for detecting  
CC altered gene expression in inbred or hybrid plants, and for screening  
CC several molecules for specific binding to the polynucleotide. The cpds  
CC are useful to identify, isolate, or extend identical or related  
CC corn-ear nucleic acid sequences from DNA libraries, and in nucleic

CC acid amplification or hybridisation techniques to follow the  
 CC expression of desirable traits through plant breeding programs.  
 CC Preferably, the cdps are used to identify, evaluate, alter, or  
 CC follow the inheritance of desired characteristics associated with  
 CC growth and development, disease resistance, environmental adaptability,  
 CC quality, and yield of corn. The cdps are also useful as molecular  
 CC markers for studying inheritance and multigene traits in a plant  
 CC breeding program. The cdps are useful for producing purified corn-ear  
 CC polypeptides by recombinant techniques. They are also useful in  
 CC diagnostic assays to detect or confirm conditions or diseases  
 CC associated with abnormal levels of cdp expression. ABX81541-ABX89140  
 CC represent corn ear-derived polynucleotides (cdps) of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site at [seqdata.uspto.gov/psipd/IDentry.html](http://seqdata.uspto.gov/psipd/IDentry.html).  
 CC  
 SQ Sequence 288 BP; 55 A; 108 C; 81 G; 44 T; 0 other;  
 Query Match 59.3%; Score 17.8; DB 25; Length 288;  
 Best Local Similarity 90.5%; Pred. No. 3.4e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 10 GTGCTACGACCTCTCTGCAGG 30  
 DB 208 GTGCTACGACCTCTCTGCAGG 228  
 RESULT 15  
 ABL72764/C  
 ID ABL72764 standard; cDNA; 289 BP.  
 AC ABL72764;  
 XX  
 DT 14-MAY-2002 (first entry)  
 XX  
 DE Corn tassels-derived polynucleotide (cdps) SEQ ID NO:2138.  
 XX  
 XX Corn; corn tassels-derived polynucleotide; cdps; hybrid breeding; CDPs;  
 KW inheritance; characteristic; growth; development; disease resistance;  
 KW environmental adaptability; quality; yield; molecular marker;  
 KW multigene trait; plant breeding; corn tassels; gene; ss.  
 XX  
 OS Zea mays.  
 XX  
 PN US2001051335-A1.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 16-APR-1999; 99US-0294093.  
 XX  
 PR 21-APR-1998; 98US-082567P.  
 XX  
 PA (LALG/) LALGUDI R V.  
 PA (ITOL/) ITO L Y.  
 PA (SHER/) SHERMAN B K.  
 XX  
 PI Lalgudi RV, Ito LY, Sherman BK;  
 XX  
 DR WPI; 2002-163647/21.  
 XX  
 PT Novel purified corn tassels-derived polynucleotide useful for  
 PT determining altered gene expression, to recover regulatory elements and  
 PT to follow inheritance of desirable characteristics through hybrid  
 PT breeding programs -  
 XX  
 PS Claim 1; SEQ ID 2138; 201pp; English.  
 XX  
 CC The present sequence describes a purified corn tassels-derived  
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
 CC encode corn tassels-derived polypeptides (CDPs). The cdps sequences (I)  
 CC can be used for determining altered gene expression, to recover  
 CC regulatory elements and to follow inheritance of desirable

CC characteristics through hybrid breeding programs. (I) are also useful  
 CC in the evaluation, and alteration of desired characteristics associated  
 CC with growth and development, disease resistance, environmental  
 CC adaptability, quality and yield, and as molecular markers for studying  
 CC inheritance of multigene traits in a plant breeding program. (I) can be  
 CC used to produce a tassels-specific profile of gene transcription, a  
 CC transcript image, to clone regulatory elements for use in transformation  
 CC vectors, to express a polypeptide, to identify, isolate or extend  
 CC identical or related corn tassels nucleic acid sequences from DNA  
 CC libraries, in nucleic acid hybridisation or amplification technologies,  
 CC as query sequences to determine homology of known sequences, as probe  
 CC for use in Southern or Northern hybridisation, and to identify the  
 CC presence of and/or to determine the degree of similarity between two  
 CC (or more) nucleic acid sequences.  
 CC  
 SQ Sequence 289 BP; 79 A; 51 C; 75 G; 84 T; 0 other;  
 Query Match 59.3%; Score 17.8; DB 24; Length 289;  
 Best Local Similarity 75.9%; Pred. No. 3.4e+02;  
 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 CCTGCGTAGTGCTACGACCTCTCTGCAGG 30  
 DB 227 CCTATATAGTAGTATGACCACTCTGCAGG 199

Search completed: September 16, 2003, 18:37:05  
 Job time : 157.297 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 18:24:35 ; Search time 38.3168 Seconds  
(without alignments)  
345.579 Million cell updates/sec

Title: US-09-594-065-3

Perfect score: 30  
Sequence: 1 CCCGCGTAGTGTGACGACCTCCTGCAGG 30

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.4	64.7	483	US-09-252-991A-6704	Sequence 6704, Ap
C 2	19.4	64.7	510	US-09-252-991A-6599	Sequence 6599, Ap
C 3	19.4	64.7	3084	US-09-252-991A-6539	Sequence 6539, Ap
C 4	19.4	64.7	3273	US-09-252-991A-6578	Sequence 6578, Ap
C 5	18.8	62.7	1368	US-08-738-172-3	Sequence 3, Appli
C 6	18.8	62.7	2936	US-08-738-172-1	Sequence 1, Appli
C 7	18.8	62.7	4454	US-08-738-172-2	Sequence 2, Appli
C 8	17.8	59.3	288	US-09-313-294A-613	Sequence 613, App
C 9	17.8	59.3	1766	US-09-904-615-29	Sequence 29, Appl
C 10	17.8	59.3	2035	PCT-US93-12464-1	Sequence 1, Appli
C 11	17.8	59.3	12848	US-09-453-702B-252	Sequence 252, App
C 12	17.8	59.3	48908	US-09-453-702B-137	Sequence 137, App
C 13	17.4	58.0	1520	US-09-620-312D-458	Sequence 458, App
C 14	17.2	57.3	8460	US-08-469-005A-9	Sequence 9, Appli
C 15	17.2	57.3	8519	US-09-261-907-1	Sequence 1, Appli
C 16	17.2	57.3	11725	US-08-756-506-1	Sequence 1, Appli
C 17	17.2	57.3	49377	US-08-764-233A-1	Sequence 14975, A
C 18	17.2	56.7	1803	US-09-252-991A-14975	Sequence 14842, A
C 19	17	56.7	2076	US-09-252-991A-14842	Sequence 913, App
C 20	17	56.7	4530	US-09-221-017B-913	Sequence 460, App
C 21	16.8	56.0	100	US-09-060-756-460	Sequence 460, App
C 22	16.8	56.0	100	US-09-060-756-460	Sequence 460, App
C 23	16.8	56.0	1214	US-09-506-286B-15	Sequence 18, Appl
C 24	16.8	56.0	1214	US-09-506-286B-15	Sequence 15, Appl
C 25	16.8	56.0	1214	US-09-762-861B-15	Sequence 18, Appl
C 26	16.8	56.0	1214	US-09-762-861B-18	Sequence 18, Appl
C 27	16.8	56.0	1230	US-09-252-991A-7646	Sequence 7646, Ap

28	16.8	56.0	1241	4	US-09-506-286B-13	Sequence 13, Appl
29	16.8	56.0	1241	4	US-09-506-286B-16	Sequence 16, Appl
30	16.8	56.0	1241	4	US-09-762-861B-13	Sequence 13, Appl
31	16.8	56.0	1241	4	US-09-762-861B-16	Sequence 16, Appl
C 32	16.8	56.0	1270	4	US-09-016-434-756	Sequence 756, App
C 33	16.8	56.0	1445	4	US-09-697-367-19	Sequence 19, Appl
C 34	16.8	56.0	2104	1	US-08-592-126-96	Sequence 96, Appl
C 35	16.8	56.0	2104	2	US-08-633-879C-1	Sequence 96, Appl
C 36	16.8	56.0	2194	2	US-08-633-879C-1	Sequence 96, Appl
C 37	16.8	56.0	2277	4	US-09-506-286B-46	Sequence 46, Appl
C 38	16.8	56.0	2277	4	US-09-506-286B-49	Sequence 49, Appl
C 39	16.8	56.0	2341	4	US-09-506-286B-44	Sequence 44, Appl
C 40	16.8	56.0	2341	4	US-09-506-286B-47	Sequence 47, Appl
C 41	16.8	56.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli
C 42	16.4	54.7	432	4	US-09-397-787-191	Sequence 191, Appl
C 43	16.4	54.7	447	4	US-09-183-861-78	Sequence 78, Appl
C 44	16.4	54.7	447	4	US-09-022-765-78	Sequence 78, Appl
C 45	16.4	54.7	447	4	US-09-551-974A-78	Sequence 78, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-252-991A-6704/c
; Sequence 6704, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6704
; LENGTH: 483
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6704

Query Match          64.7% Score 19.4; DB 4; Length 483;
Best Local Similarity 79.3%; Pred. No. 10;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCCGCGTAGTGTGACGACCTCCTGCAGG 29
Db 35 CCCGCGTAGTGTGACGACCTCCTGCAGG 7

RESULT 2
US-09-252-991A-6599
; Sequence 6599, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6599
; LENGTH: 510
; TYPE: DNA
```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,172
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,965
FILING DATE: October 27, 1995
APPLICATION NUMBER: 08/465,003
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/575,354
FILING DATE: December 20, 1995
APPLICATION NUMBER: 08/240,158
FILING DATE: May 10, 1994
APPLICATION NUMBER: 08/229,329
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/158
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2936 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-738-172-1

Query Match          62.7%; Score 18.8; DB 2; Length 2936;
Best Local Similarity 76.7%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CCTGCGTAGTGGTACGACCTCTCCAGGG 30
DB 1351 CCGTGGTACTGCTTCGGGCTCCAGCCGGG 1322

RESULT 7
US-08-738-172-2/c
Sequence 2, Application US/08738172
Patent No. 5939257
GENERAL INFORMATION:
APPLICANT: Szaaz, Joseph
APPLICANT: Davis, Maria
TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
```

```

MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,172
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/005,965
FILING DATE: October 27, 1995
APPLICATION NUMBER: 08/465,003
FILING DATE: June 5, 1995
APPLICATION NUMBER: 08/575,354
FILING DATE: December 20, 1995
APPLICATION NUMBER: 08/240,158
FILING DATE: May 10, 1994
APPLICATION NUMBER: 08/229,329
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/158
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4454 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-738-172-2

Query Match          62.7%; Score 18.8; DB 2; Length 4454;
Best Local Similarity 76.7%; Pred. No. 24;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
```

```

QY 1 CCGTGGTACTGCTTCGGGCTCCAGCCGGG 30
DB 1738 CCGTGGTACTGCTTCGGGCTCCAGCCGGG 1709

RESULT 8
US-09-313-294A-613
Sequence 613, Application US/09313294A
Patent No. 6476212
GENERAL INFORMATION:
APPLICANT: Ialagudi, Raghunath V.
APPLICANT: Ito, Laura Y.
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
FILE REFERENCE: PL-0017 US
CURRENT APPLICATION NUMBER: US/09/313,294A
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 7600
SOFTWARE: PERL Program
SEQ ID NO 613
LENGTH: 288
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. 6476212 700549530H1
US-09-313-294A-613
```

```

Query Match          59.3%; Score 17.8; DB 4; Length 288;
Best Local Similarity 90.5%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GTGTCAGACCTCTCCAGGG 30
```

Db 208 GTGTACGACTCCGGCTGAG 228

## RESULT 9

US-09-904-615-29/c  
Sequence 29, Application US/09904615

Patent No. 6566325  
GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: 49 Human Secreted Proteins

FILE REFERENCE: P2032P1

CURRENT APPLICATION NUMBER: US/09/904,615

CURRENT FILING DATE: 2001-07-16

PRIOR APPLICATION NUMBER: 09/511,554

PRIOR FILING DATE: 2000-02-23

PRIOR APPLICATION NUMBER: 60/097,917

PRIOR FILING DATE: 1998-08-25

PRIOR APPLICATION NUMBER: 60/098,634

PRIOR FILING DATE: 1998-08-31

NUMBER OF SEQ ID NOS: 170

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 29

LENGTH: 1766

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (1743)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (1748)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (1749)

OTHER INFORMATION: n equals a,t,g, or c

US-09-904-615-29

Query Match

Best Local Similarity 59.3%; Score 17.8; DB 4; Length 1766;

Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCTCGGTAGTGTGACGACTCTCTGAG 29

Db 1312 CCTCGGTGTGCTCTCCACCTCTGAG 1284

RESULT 10  
PCT-US93-12464-1  
Sequence 1, Application PC/TUS9312464

GENERAL INFORMATION:

APPLICANT: The Regents of the University of California, et al.

TITLE OF INVENTION: METHOD FOR INHIBITION OF CELL

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ

STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR

CITY: LOS ANGELES

STATE: CALIFORNIA

COUNTRY: US

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/12464

FILING DATE: 21-DEC-1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: HOWELLS, STACY L.

REGISTRATION NUMBER: 34,842

REFERENCE/DOCKET NUMBER: PD2929

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 455-5100

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 1;

SEQUENCE CHARACTERISTICS:

LENGTH: 2035 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

IMMEDIATE SOURCE:

CLONE: ALPHA-L-FUCOSIDASE

FEATURE:

NAME/KEY: CDS

LOCATION: 19..1401

PCT-US93-12464-1

Query Match

Best Local Similarity 59.3%; Score 17.8; DB 5; Length 2035;

Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCTCGGTAGTGTGACGACTCTCTGAG 29

Db 351 CCGGAGGAGTGGCCGACCTCTTCAG 379

RESULT 11  
US-09-453-702B-252/c

Sequence 252, Application US/09453702B

Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blatner, Frederick R.

Burland, Valerie

Perna, Nicole T.

Plunkett, Guy

Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Plunkney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296,95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 252:

SEQUENCE CHARACTERISTICS:

LENGTH: 12848

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 252:



US-09-453-702B-252

Query Match 59.3%; Score 17.8; DB 4; Length 12848;  
Best Local Similarity 75.9%; Pred. No. 74;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCTGCTAGTGGTACGACCTCTCGACGG 30  
Db 4539 CTTCCTAGCGGAGGAACTCTCGACGG 4511

RESULT 12

US-09-453-702B-137/c

Sequence 137, Application US/09453702B

Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blatner, Frederick R.

Burland, Valerie

Perna, Nicole T.

Plunkett, Guy

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles &amp; Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-Dec-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296,95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 137:

SEQUENCE CHARACTERISTICS:

LENGTH: 48908

TYPE: nucleic acid

STRADEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 137:

US-09-453-702B-137

Query Match 59.3%; Score 17.8; DB 4; Length 48908;

Best Local Similarity 75.9%; Pred. No. 86;

Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 2 CCTGCTAGTGGTACGACCTCTCGACGG 30

Db 15720 CTTCCTAGCGGAGGAACTCTCGACGG 15692

RESULT 13

US-09-620-312D-458

Sequence 458, Application US/09620312D

Patent No. 6569862

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyun

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Yidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yundong

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: John Tillinghast

APPLICANT: Dmanac, Radoje T.

TITLE OF INVENTION: No. 656962el Nucleic Acids and

FILE REFERENCE: Polypeptides

CURRENT APPLICATION NUMBER: US/09/620,312D

PRIOR FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: pf\_fl\_genes Version 1.0

SEQ ID NO 458

LENGTH: 1520

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (271)..(1161)

US-09-620-312D-458

Query Match 58.0%; Score 17.4; DB 4; Length 1520;

Best Local Similarity 77.8%; Pred. No. 87;

Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 CCTGCTAGTGGTACGACCTCTCGACG 28

Db 277 CCGGAGTCTGTGACGACCTCTCGACG 303

RESULT 14

US-08-469-005A-9/c

Sequence 9, Application US/08469005A

Patent No. 5665874

GENERAL INFORMATION:

APPLICANT: KUHADJA, FRANCIS P.

APPLICANT: PASTERNAK, GARY A.

TITLE OF INVENTION: CANCER RELATED ANTIGEN

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: BAKER &amp; BOTT, L.L.P.

STREET: 1299 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20004-2400

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,005A

FILING DATE: 05-JUN-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/188,426

```

/ FILING DATE: 24-JAN-1994
/ APPLICATION NUMBER: 08/096,908
/ FILING DATE: 26-JUL-1993
/ APPLICATION NUMBER: 07/917,716
/ FILING DATE: 24-JUL-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Posortake, Laurence H
/ REGISTRATION NUMBER: 34,698
/ REFERENCE/DOCKET NUMBER: 062482-0113
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-639-7700
/ TELEFAX: 202-639-7890
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8460 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE:
/ ORIGINAL SOURCE:
/ FEATURE:
/ NAME/KEY: Coding Sequence
/ LOCATION: 124...7650
/ OTHER INFORMATION:
US-08-469-005A-9

```

```

Query Match      57.3%; Score 17.2; DB 1; Length 8460;
Best Local Similarity 73.3%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY      1 CCCTGCGTAGTGTAGCACTCCTCGAGG 30
DB      3550 CCTTGCACAGTGCAGCTCCTCCTCGAGG 3521

```

```

RESULT 15
US-09-261-907-1/C
/ Sequence 1, Application US/09261907A
/ Patent No. 6294364
/ GENERAL INFORMATION:
/ APPLICANT: ELLIS, CATHERINE
/ APPLICANT: LONSDALE, JOHN
/ APPLICANT: BERGSMAN, DEBK U.
/ APPLICANT: MOONEY, JEFFREY L.
/ APPLICANT: DEPIERA, MEGAN E.
/ APPLICANT: CHAPMAN, CONRAD
/ TITLE OF INVENTION: HUMAN PAs
/ FILE REFERENCE: GP-70603
/ CURRENT APPLICATION NUMBER: US/09/261,907A
/ CURRENT FILING DATE: 1999-03-03
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 1
/ LENGTH: 8519
/ TYPE: DNA
/ ORGANISM: HOMO SAPIENS
US-09-261-907-1

```

```

Query Match      57.3%; Score 17.2; DB 3; Length 8519;
Best Local Similarity 73.3%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY      1 CCCTGCGTAGTGTAGCACTCCTCGAGG 30
DB      3582 CCTTGCACAGTGCAGCTCCTCCTCGAGG 3553

```

```

Search completed: September 16, 2003, 20:42:38
Job time : 40.3168 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 19:14:25 ; Search time 108.416 Seconds  
(without alignments)  
680.690 Million cell updates/sec

Title: US-09-594-065-3

Perfect score: 30

Sequence: 1 cccgcgtagtgcgtacgacctcgcaggg 30

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1660708 seqs, 122959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCF\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/FCIS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	64.7	1323	14 US-10-156-761-4750	Sequence 4750, Ap
2	19.4	64.7	110079	14 US-10-175-523-96	Sequence 96, Appl
3	19.4	64.7	9025608	14 US-10-156-761-1	Sequence 1, Appl
4	18.8	62.7	576	10 US-09-738-626-2147	Sequence 2147, Ap
5	18.8	62.7	1368	9 US-09-191-989-3	Sequence 3, Appl
6	18.8	62.7	1593	14 US-10-156-761-5071	Sequence 5071, Ap
7	18.8	62.7	2936	9 US-09-191-989-1	Sequence 1, Appl
8	18.8	62.7	4434	9 US-09-191-989-2	Sequence 2, Appl
9	18.8	62.7	99916	10 US-09-816-095-3	Sequence 3, Appl
10	18.4	62.7	3309400	10 US-09-738-626-1	Sequence 39424, A
11	18.4	61.3	536	13 US-10-027-632-39424	Sequence 39424, A
12	18.4	61.3	536	13 US-10-027-632-58202	Sequence 58202, A
13	18.4	61.3	536	13 US-10-027-632-58209	Sequence 58209, A
14	18.4	61.3	536	13 US-10-027-632-309173	Sequence 309173, A
15	18.4	61.3	4705	12 US-10-312-918-1	Sequence 1, Appl
16	18.4	61.3	4786	12 US-10-312-918-3	Sequence 3, Appl

C 17	18.4	61.3	1503841	9 US-09-795-668-1	Sequence 1, Appl
C 18	18.4	61.3	1503841	9 US-09-795-668-1	Sequence 1, Appl
C 19	18.4	61.3	1503841	10 US-09-946-807-1	Sequence 1, Appl
C 20	17.8	59.3	289	9 US-09-294-093B-2138	Sequence 2138, Ap
C 21	17.8	59.3	630	14 US-10-156-761-7444	Sequence 7444, Ap
C 22	17.8	59.3	745	13 US-10-027-632-146355	Sequence 146355, Ap
C 23	17.8	59.3	1766	9 US-09-739-254-29	Sequence 29, Appl
C 24	17.8	59.3	1766	9 US-09-904-615-29	Sequence 29, Appl
C 25	17.8	59.3	1766	12 US-10-055-098-29	Sequence 29, Appl
C 26	17.8	59.3	1766	14 US-10-054-988-29	Sequence 29, Appl
C 27	17.8	59.3	2035	12 US-09-873-367C-711	Sequence 711, App
C 28	17.8	59.3	2032	14 US-10-156-761-4847	Sequence 4847, Ap
C 29	17.8	59.3	2162	14 US-10-158-846-13433	Sequence 13433, A
C 30	17.8	59.3	12848	12 US-10-205-194-30	Sequence 30, Appl
C 31	17.8	59.3	12848	14 US-10-114-170-252	Sequence 252, App
C 32	17.8	59.3	48908	14 US-10-114-170-137	Sequence 137, App
C 33	17.8	59.3	9025608	14 US-10-156-761-1	Sequence 1, Appl
C 34	17.6	58.7	4829	12 US-10-021-660-36	Sequence 36, App
C 35	17.6	58.7	12404	13 US-10-108-605-236	Sequence 236, App
C 36	17.4	58.0	81	10 US-09-783-590-7470	Sequence 7470, Ap
C 37	17.4	58.0	403	11 US-09-918-995-16763	Sequence 16763, A
C 38	17.4	58.0	606	10 US-09-951-470-6	Sequence 6, Appl
C 39	17.4	58.0	700	13 US-10-027-632-150978	Sequence 150978, A
C 40	17.4	58.0	700	13 US-10-027-632-150979	Sequence 150979, A
C 41	17.4	58.0	700	13 US-10-027-632-150980	Sequence 150980, A
C 42	17.4	58.0	799	9 US-09-764-870-263	Sequence 263, App
C 43	17.4	58.0	799	14 US-10-125-540-263	Sequence 263, App
C 44	17.4	58.0	879	10 US-09-951-470-4	Sequence 4, Appl
C 45	17.4	58.0	1520	14 US-10-037-270-458	Sequence 458, App

## ALIGNMENTS

RESULT 1  
US-10-156-761-4750  
Sequence 4750, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIORITY APPLICATION NUMBER: JP 2001-204089  
PRIORITY FILING DATE: 2001-05-30  
PRIORITY APPLICATION NUMBER: JP 2001-272697  
PRIORITY FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 4750  
LENGTH: 1323  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1323)  
US-10-156-761-4750

Query Match 64.7%; Score 19.4; DB 14; Length 1323;  
Best Local Similarity 79.3%; Pred. No. 27;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 cccgcgtagtgcgtacgacctcgcaggg 29  
DB 669 cccgcgtagtgcgtacgacctcgcaggg 697

```
RESULT 2
US-10-175-523-96/c
; Sequence 96, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Paleyman, Michael
; APPLICANT: Rajan, Priti
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPTS)
; FILE REFERENCE: 3235/13795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 96
; LENGTH: 110079
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(110079)
; OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-10-175-523-96

Query Match      64.7%; Score 19.4; DB 14; Length 110079;
Best Local Similarity 79.3%; Pred. No. 26;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  CCCTGCGTAGTGTGATGACCTCTCTGCAGG 29
Db      4399  CACTGCTAGTGCATGTCATCTCTGCAGG 4371

RESULT 3
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
```

```
; ORGANISM: Streptomyces avermiltilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      64.7%; Score 19.4; DB 14; Length 9025608;
Best Local Similarity 79.3%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      1  CCCTGCGTAGTGTGATGACCTCTCTGCAGG 29
Db      5809596  CCCCCGGAAGCGGTACGACGTCTCTGCG 5809568

RESULT 4
US-09-738-626-2147/c
; Sequence 2147, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2147
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2147

Query Match      62.7%; Score 18.8; DB 10; Length 576;
Best Local Similarity 90.9%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      9  AGTGTACGACCTCTCTGCAGG 30
Db      258  AGTGTACGACCACTGCAAGG 237

RESULT 5
US-09-191-989-3
; Sequence 3, Application US/09191989
; Patent No. US2002005098A1
; GENERAL INFORMATION:
; APPLICANT: Szaez, Joseph
; APPLICANT: Davis, Maria
; TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
```

```

; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191.989
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/738,172
; FILING DATE: October 25, 1996
; APPLICATION NUMBER: 08/465,003
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/575,354
; FILING DATE: December 20, 1995
; APPLICATION NUMBER: 08/240,158
; FILING DATE: May 10, 1994
; APPLICATION NUMBER: 08/229,329
; FILING DATE: April 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Wardburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1368 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-191-989-3

Query Match
Best Local Similarity 62.7%; Score 18.8; DB 9; Length 1368;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTACGACCTCTGCAGGG 30
Db 197 CCGTGGCTACTGCTTCGGCCTTCACGCCGG 226

RESULT 6
US-10-156-761-5071
; Sequence 5071, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5071
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Streptomyces avermiltilis
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1593)
;
US-10-156-761-5071

Query Match
Best Local Similarity 62.7%; Score 18.8; DB 14; Length 1593;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTACGACCTCTGCAGGG 30
Db 720 CCGGCGCAGCGCTACGACATCTGCTGGG 749

RESULT 7
US-09-191-989-1/c
; Sequence 1, Application US/09191989
; Patent No. US20020055098A1
; GENERAL INFORMATION:
; APPLICANT: Szasz, Joseph
; APPLICANT: Davis, Maria
; TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191.989
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/738,172
; FILING DATE: October 25, 1996
; APPLICATION NUMBER: 08/465,003
; FILING DATE: June 5, 1995
; APPLICATION NUMBER: 08/575,354
; FILING DATE: December 20, 1995
; APPLICATION NUMBER: 08/240,158
; FILING DATE: May 10, 1994
; APPLICATION NUMBER: 08/229,329
; FILING DATE: April 18, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Wardburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 222/158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2936 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-191-989-1

Query Match
Best Local Similarity 76.7%; Score 18.8; DB 9; Length 2936;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTACGACCTCTGCAGGG 30
```

Db 1351 CCCTGCGTACTGCTCGCCTCCAGCCGGG 1322

## RESULT 8

US-09-191-989-2/c  
; Sequence 2, Application US/09191989  
; Patent No. US20020055098A1  
; GENERAL INFORMATION:  
; APPLICANT: Szasz, Joseph  
; APPLICANT: Davis, Maria  
; TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/191,989  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/738,172  
; FILING DATE: October 25, 1996  
; APPLICATION NUMBER: 08/465,003  
; FILING DATE: June 5, 1995  
; APPLICATION NUMBER: 08/575,354  
; FILING DATE: December 20, 1995  
; APPLICATION NUMBER: 08/240,158  
; FILING DATE: May 10, 1994  
; APPLICATION NUMBER: 08/229,329  
; FILING DATE: April 18, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 222/158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 673510  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4454 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

US-09-191-989-2

Query Match 62.7%; Score 18.8; DB 9; Length 4454;  
Best Local Similarity 76.7%; Pred. No. 49;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CCCTGCGTACTGCTCGCCTCCAGCCGGG 30  
Db 1738 CCCTGCGTACTGCTCGCCTCCAGCCGGG 1709

## RESULT 9

US-09-816-095-3/c  
; Sequence 3, Application US/09816095  
; Patent No. US20020137164A1  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weimin

; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES  
; FILE REFERENCE: C1001147  
; CURRENT APPLICATION NUMBER: US/09/816,095  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 3  
; LENGTH: 99916  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)-(99916)  
; OTHER INFORMATION: n = A,T,C,G

US-09-816-095-3

Query Match 62.7%; Score 18.8; DB 10; Length 99916;  
Best Local Similarity 76.7%; Pred. No. 48;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CCCTGCGTACTGCTCGCCTCCAGCCGGG 30  
Db 1748 CCCTGCGGCGCTCGCAGCCTCTCGAGGG 1719

## RESULT 10

US-09-738-626-1/c  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIRO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patentin ver. 3.0  
; SEQ ID NO: 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum

US-09-738-626-1

Query Match 62.7%; Score 18.8; DB 10; Length 3309400;  
Best Local Similarity 90.9%; Pred. No. 44;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 AGTGTACGACCTCTCTGACGGG 30  
Db 2066823 AGTGTACGACCACTCGCAGG 2066802

## RESULT 11

US-10-027-632-39424/c  
; Sequence 39424, Application US/10027632  
; GENERAL INFORMATION:

```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 39424
LENGTH: 536
TYPE: DNA
ORGANISM: Human
US-10-027-632-39424
```

```
Query Match          61.3%; Score 18.4; DB 13; Length 536;
Best Local Similarity 78.6%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 3 CTGCGTAGTGTACGACCTCTCCAGAGG 30
    ||||| | | ||||| |||||
Db 117 CTGCGTAGTGTACGACCTCTCCAGAGG 90
```

```
RESULT 12
US-10-027-632-58202/c
Sequence 58202, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 58202
LENGTH: 536
TYPE: DNA
ORGANISM: Human
US-10-027-632-58202
```

```
Query Match          61.3%; Score 18.4; DB 13; Length 536;
Best Local Similarity 78.6%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 3 CTGCGTAGTGTACGACCTCTCCAGAGG 30
    ||||| | | ||||| |||||
Db 117 CTGCGTAGTGTACGACCTCTCCAGAGG 90
```

```
RESULT 13
US-10-027-632-59209/c
Sequence 59209, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 59209
LENGTH: 536
TYPE: DNA
ORGANISM: Human
US-10-027-632-59209
```

```
Query Match          61.3%; Score 18.4; DB 13; Length 536;
Best Local Similarity 78.6%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
Qy 3 CTGCGTAGTGTACGACCTCTCCAGAGG 30
    ||||| | | ||||| |||||
Db 117 CTGCGTAGTGTACGACCTCTCCAGAGG 90
```

```
RESULT 14
US-10-027-632-309173/c
Sequence 309173, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 309173
```

; LENGTH: 536  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-309173

Query Match 61.3%; Score 18.4; DB 13; Length 536;  
 Best Local Similarity 78.6%; Pred. No. 76;  
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 CTGCGTAGTGATGACGACCTCTGCGAGG 30  
 DB 117 CTGCGTAGTGACGACCTCTGCGAGG 90

# RESULT 15

US-10-312-918-1  
 ; Sequence 1, Application US/10312918  
 ; Publication No. US20030162740A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Regents of the University of California  
 ; APPLICANT: Wang, Jean Y.J.  
 ; APPLICANT: Vigneri, Paolo  
 ; TITLE OF INVENTION: New Strategy for Leukemia Therapy  
 ; FILE REFERENCE: 6627-PC0159  
 ; CURRENT APPLICATION NUMBER: US/10/312,918  
 ; CURRENT FILING DATE: 2002-12-27  
 ; PRIOR APPLICATION NUMBER: 60/215,595  
 ; PRIOR FILING DATE: 2000-06-30  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4705  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien/ Mus musculus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(4704)  
 ; OTHER INFORMATION:  
 US-10-312-918-1

Query Match 61.3%; Score 18.4; DB 12; Length 4705;  
 Best Local Similarity 78.6%; Pred. No. 75;  
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CCTGCGTAGTGATGACGACCTCTGCGAGG 29  
 DB 3414 CCTGCGTAGTGATGACGACCTCTGCGAGG 3441

Search completed: September 16, 2003, 22:54:38  
 Job time : 131.416 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:22:25 / Search time 1518.42 Seconds  
(without alignments)  
480.194 Million cell updates/sec

Title: US-09-594-065-3

Perfect score: 30

Sequence: 1 CCCGCGTGTGTCGACGACCTCCGCGAGG 30

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_estchum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_pro:\*  
25: em\_ges\_rtd:\*  
26: em\_ges\_phg:\*  
27: em\_ges\_vrl:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	22.2	74.0	900	10	BG288880	602383822
2	21	70.0	993	14	BY707129	BY707129
3	20.4	66.0	337	13	BX264906	BX264906
4	20.4	66.0	516	12	BM864032	mgcm008xi

Result No.	Score	Match	Length	DB	ID	Description
5	20.4	68.0	880	29	CNS04LYM	AL296743 Tetradon
6	20	66.7	752	9	AA896672	AA896672 vx54d07.r
7	20	66.7	723	12	B1327791	B1327791 602979537
8	19.8	66.0	450	11	BY463661	BY463661
9	19.6	65.3	161	14	CD475739	CD475739 na003.15m
10	19.4	64.7	380	10	BB872020	BB872020 BB872020
11	19.4	64.7	406	9	AA709478	AA709478 vv34d05.r
12	19.4	64.7	432	13	BY680656	BY680656 BY680656
13	19.4	64.7	478	14	CB790399	CB790399 AMGNNUC:N
14	19.4	64.7	434	10	BB857961	BB857961 BB857961
15	19.4	64.7	531	10	BF022130	BF022130 uy47b10.y
16	19.4	64.7	576	12	B1688644	B1688644 60311128
17	19.4	64.7	623	28	A2634235	A2634235 1M0489020
18	19.4	64.7	633	10	BE269697	BE269697 601185944
19	19.4	64.7	720	10	BE298602	BE298602 602396992
20	19.4	64.7	724	12	B1692432	B1692432 603342935
21	19.4	64.7	884	14	CA977558	CA977558 AGENCOURT
22	19.4	64.7	937	10	BF106954	BF106954 601822889
23	19.4	64.7	997	29	CNS040LD	AL269050 Tetradon
24	19.4	64.7	1160	29	CC254336	CC254336 CH261-63F
25	19.4	64.7	2225	11	AK007059	AK007059 Mus muscu
26	19.4	64.7	2609	11	BC005727	BC005727 Mus muscu
27	19.4	64.7	2869	11	AK042427	AK042427 Mus muscu
28	19.2	64.0	380	14	CA690063	CA690063 wlm96.pk0
29	19.2	64.0	402	14	CA669616	CA669616 wlm91.pk0
30	19.2	64.0	555	14	CA666652	CA666652 wlm91.pk0
31	19.2	64.0	584	14	CA666706	CA666706 wlm91.pk0
32	19	63.3	421	10	BB261816	BB261816
33	19	63.3	432	28	AQ216405	AQ216405 HS_2262_A
34	19	63.3	568	9	AA469480	AA469480 hd18a09.x
35	19	63.3	623	10	BF252994	BF252994 EST445489
36	19	63.3	629	10	BE305546	BE305546 601102085
37	19	63.3	794	19	AL239208	AL239208 Tetradon
38	19	63.3	814	29	CNS030AP	AL252898 Tetradon
39	19	63.3	840	29	CNS04FIP	AL288394 Tetradon
40	19	63.3	1105	12	BM806374	BM806374 AGENCOURT
41	18.8	62.7	300	9	AL838394	AL838394
42	18.8	62.7	340	13	BY335529	BY335529 BY335529
43	18.8	62.7	366	2	HSW090445	Bx497096 Homo sapi
44	18.8	62.7	391	14	BY702686	BY702686
45	18.8	62.7	476	14	CB949437	CB949437 AGENCOURT

## ALIGNMENTS

RESULT 1  
LOCUS BG288880  
DEFINITION 602383822P1 NIH\_MGC\_93 Homo sapiens cDNA clone IMAGE:4512917 5', mRNA sequence.  
ACCESSION BG288880  
VERSION BG288880.1 GI:13044162  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 900)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contract: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:  
http://image.llnl.gov  
Plate: LLM10398 row: 1 column: 06  
High quality sequence stop: 718.



```

REFERENCE 1 (bases 1 to 516)
AUTHORS Ebbole,D.J., Yuan,J., Thomas,T.L., Bobrowicz,P., Lu,G., Bhatterai
TITLE 'K. and Dean,R.A.
Expressed sequence tags from the rice blast fungus, Magnaporthe
grisea
JOURNAL Unpublished
COMMENT On Mar 7, 2002 this sequence version replaced gi:19231714.
Contact: Ebbole DJ
Department of Plant Pathology & Microbiology
Texas A&M University
Peterson Bldg, MS2132, College Station, TX 77843-2132, USA
Tel: 979 845 4831
Fax: 979 845 6483
Email: d.ebbole@tamu.edu
Chromatogram file of this sequence is available, see contact person
/best nr hit (April, 22, 2003) sp|P15710|PHO4_NECCR
PHOSPHATE-REPRESSIBLE PHOSPHATE PERMEASE >g... 53 2e-06
PCR Primers
FORWARD: T3 primer
BACKWARD: T7 primer
Plate: mgcm008 row: 1 column: 16
Seq primer: T3.
Location/Qualifiers
1..516
/organism="Magnaporthe grisea"
/mol_type="mRNA"
/strain="Guy11"
/db_xref="taxon:148305"
/cdoe="mgcm008x116"
/sex="Mat-2 hermaphrodite"
/cell_type="mycelium"
/clone_id="Magnaporthe grisea CM Uni-2ap XR library"
/notes="Vector: pBluescriptSK-; Site 1: EcoRI; Site 2: XhoI
; Unidirectional cloning. EcoRI site has T3 primer and
predominantly 5' reads. T7 primer on XhoI side of insert.
Strain inoculated into complete medium grown for two days
at room temperature, 150 rpm, harvested, blended,
re-inoculated into complete medium 24 h, room temperature,
150 rpm. Sequences were processed by one of two methods.
Where a full-length alignment to the M. grisea genome
sequence was available, the EST sequence was trimmed
according to the alignment, otherwise sequence quality was
assessed using phredphrap version 991015 and trimmed
according to phd files (0.05) and for vector seqs."
BASE COUNT 97 a 193 c 147 g 79 t
ORIGIN
Query Match 68.0%; Score 20.4; DB 12; Length 516;
Best Local Similarity 80.0%; Pred. No. 9.9e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 CCCGCGCTAGTGGTACGACCTCCGCGAGGG 30
| ||| ||||| ||||| ||||| |||||
Db 121 CACTGTGAGTGGTACGACTTCCTCCAGGG 150
RESULT 5
CNS04LYM/c CNS04LYM 880 bp DNA linear GSS 01-SEP-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION 119008 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL2966743
VERSION AL2966743.1 GI:8035323
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis
ORGANISM Tetradon nigroviridis
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetradon.
REFERENCE 1
AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

```

Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,  
 Sautin, W., and Weissenbach, J.  
 Estimate of human gene number provided by genome-wide analysis  
 using Tetradon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL MEDLINE 20296633  
 PUBMED 10835645  
 REFERENCE 2  
 AUTHORS Roest Collin, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C.,  
 Fizames, C., Rischer, C., Bouneau, L., Billault, A., Quetier, F.,  
 Sautin, W., Bernot, A., and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)

JOURNAL MEDLINE 20359837  
 PUBMED 10899143  
 REFERENCE 3  
 AUTHORS (baees 1 to 880)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetradon nigroviridis  
 genome. For more information, please take a look at  
 http://www.genoscope.cns.fr/Tetradon.  
 Location/Qualifiers  
 1.880  
 /organism="Tetradon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="119008"  
 /clone\_1lb="G"  
 /note="Genoscope sequence ID : COB0119BH04SPL-end :  
 PTC-01"

BASE COUNT 168 a 250 c 256 g 184 t 22 others  
 ORIGIN

Query Match 68.0%; Score 20.4; DB 29; Length 880;  
 Best Local Similarity 80.0%; Pred. No. 1.1e+03;  
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCCTGCGTAGTGTACGACCTCTCGAGG 30  
 |||||  
 DB 485 CGCTGCGTAGTGTGCGCAGCTCTCGAGG 456

RESULT 6  
 AA896672 652 bp mRNA linear EST 06-APR-1998  
 LOCUS vx54d07.r1 StrataGene mouse macrophage (#937306) Mus musculus cDNA  
 DEFINITION clone IMAGE:1279021.5 similar to gb:54156.rnal CELLULAR TUMOR  
 ANTIGEN P53 (HUMAN); gb:X01237 Mouse mRNA for cellular tumor  
 antigen p53 (MOUSE); mRNA sequence.  
 ACCESSION AA896672  
 VERSION AA896672.1 GI:3033065  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (baees 1 to 652)  
 Mairra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Scheinberg, K., Stepcie, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 The WashU-HMI Mouse EST Project  
 Unpublished  
 Contact: Mairra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:670821  
 Seq primer: -28m3 rev1 ET from Amerham  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1.652  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1279021"  
 /tissue\_type="macrophage"  
 /dev\_stage="WEHI-3 cell line"  
 /lab\_host="GOLR (kanamycin resistant)"  
 /clone\_1lb="Stratagene mouse macrophage (#937306)"  
 /note="Organ: blood; Vector: pBlueScript SK-; Site 1:  
 EcorI; Site 2: XhoI; Cloned unidirectionally. Primer:  
 oligo dt. WEHI-3 cell line. Average insert size: 1.5 kb;  
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG  
 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 136 a 191 c 176 g 148 t 1 others  
 ORIGIN

Query Match 66.7%; Score 20; DB 9; Length 652;  
 Best Local Similarity 82.1%; Pred. No. 1.4e+03;  
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CCCTGCGTAGTGTACGACCTCTCGAG 28  
 |||||  
 DB 276 CCCTGCGTAGTGTGAGAGACTCTCGAG 303

RESULT 7  
 B1327791 723 bp mRNA linear EST 30-JUL-2001  
 LOCUS B1327791/c 602979537F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5132624 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1327791  
 VERSION B1327791.1 GI:15012448  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (baees 1 to 723)  
 NIH-MGC http://mgc.nci.nih.gov/  
 NIH-MGC http://mgc.nci.nih.gov/  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1M11325 row: 1 column: 09  
 High quality sequence start: 2  
 High quality sequence stop: 609.  
 Location/Qualifiers  
 1.723  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5132624"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_1lb="NCI\_CGAP\_L19"  
 /note="Organ: liver; Vector: PCMV-SPORT6; Site\_1: NciI;

Site\_2: Sal1; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 Kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 197 a 176 c 211 g 139 t  
ORIGIN  
Query Match 66.7%; Score 20; DB 12; Length 723;  
Best Local Similarity 82.1%; Pred. No. 1.5e+03;  
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 CCCGCGTAGTGTGACCTCTCTGAG 28  
718 CCATCGGAGCTGTGACCTCTCTGAG 691

RESULT 8 BY463661 450 bp mRNA linear EST 13-DEC-2002  
LOCUS BY463661 RIKEN full-length enriched, 10 days neonate heart Mus  
DEFINITION musculus cDNA clone K630144J20 3', mRNA sequence.  
ACCESSION BY463661  
VERSION BY463661.1 GI:26765911  
KEYWORDS EST  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Oatono, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,  
Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,  
Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chochia, C., Corbani,  
L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,  
A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
Gough, J., Grimmond, S., Gustlich, S., Hirokawa, N., Jackson, I. J.,  
Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziarski, R. M.,  
King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,  
P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
H., Nagashima, T., Numata, K., Okido, T., Pavan, M. J., Perte, G.,  
Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,  
Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,  
B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semp, C. A., Setou,  
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,  
R. D., Tomita, M., Verardo, R., Wagner, L., Wahlested, C., Wang, Y.,  
Watanabe, Y., Wells, C., Wilming, L. G., Wyszynski, B. A., Yanagisawa,  
M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura,  
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Shuhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,  
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multichannel sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

FEATURES  
source

1..450  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K630144J20"  
/tissue\_type="heart"  
/dev\_stage="10 days neonate"  
/clone\_lib="RIKEN full-length enriched, 10 days neonate  
heart"  
BASE COUNT 102 a 118 c 124 g 104 t 2 others

Query Match 66.0%; Score 19.8; DB 13; Length 450;  
Best Local Similarity 87.5%; Pred. No. 1.6e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

7 GTAGTGTACGACCTCTCTGACGG 30  
303 GGAGTGTCTTCACACTCTCTGACGG 326

RESULT 9 CD475739 161 bp mRNA linear EST 04-JUN-2003  
LOCUS CD475739/c  
DEFINITION nad03-15me2-e07 Nad03 Nuphar advena cDNA clone nad03-15me2-e07 5',  
mRNA sequence.  
ACCESSION CD475739  
VERSION CD475739.1 GI:31397007  
KEYWORDS EST.  
SOURCE Nuphar advena  
ORGANISM Nuphar advena  
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; stem Magnoliophyta; Nymphaeaceae;  
Nuphar.

1 (bases 1 to 161)  
depamphilis, C., Carlson, J., Ma, H., Frohlich, M., Tanksey, S.,  
leebens-Mack, J., Field, D., Arrington, J., Zahn, L., Kong, H.,  
Druckemiller, M., Landherr, L., Hu, Y., Ilut, D., Wall, K., Plock,  
S., Chioresu, S., Albert, V., Doyle, J., Miller, W., Oppenheimer, D.,  
Soltis, D., Soltis, P. and Theissen, G.

Generation of ESTs from early flower buds of Nuphar advena  
Unpublished  
Title  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Contact: Claude depamphilis or James leebens-Mack  
Mueller Laboratory  
Penn State University  
208 Mueller Laboratory, Department of Biology, ATTN Rm212, Penn  
State University, University Park, PA 16802, USA  
Tel: 814 863 6413  
Fax: 814 865 9131  
Email: cw33@psu.edu or jhl10@psu.edu  
The sequence provided is trimmed of vector and low quality regions.  
Full sequence and original trace file are available from the Plant

Genome Network website (<http://pgrn.cornell.edu>)  
 Place: nad03-15ms2 row: e column: 07  
 Seq primer: M13F.  
 Location/Qualifiers

FEATURES  
 source

1.361  
 /organism="Nuphar advena"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:77108"  
 /clone="nad03-15ms2-e07"  
 /rname\_type="flower buds"  
 /dev\_stage="c=2.5mm"  
 /lab\_host="SOLR"  
 /clone\_lib="Nad03"

/note="Vector: pBluescript SK (+/-); site 1: EcoRI; site 2: XhoI; Only floral buds with diameter of 2.5 mm of less were used for RNA isolation. This is a directionally cloned, non-normalized library. Avg. insert length: 1134; Primers: M13F and M13R; Antibiotic: 50 ug/ml Ampicillin; Primary titer: 286 pfu total; Amplified titer: 3.2E10 pfu/ml; Mass Excised Titer: 5E10 total; This library has been generated by the Floral Genome Project (FGP). The Floral Genome Project is funded by NSF's Plant Genome Research Program (DBI-0115684). More information about the project can be obtained at <http://fgp.bio.psu.edu>"

BASE COUNT  
 ORIGIN

Query Match 65.3%; Score 19.6; DB 14; Length 161;  
 Best Local Similarity 84.6%; Pred. No. 1.7e+03;  
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCCTAGTGATGACCTCTGCAGG 30  
 DB 110 GCCTAGTGACGACCTCTGCAGG 85

RESULT 10 380 bp mRNA linear EST 27-NOV-2001  
 BB872020 RIKEN full-length enriched, 16 days neonate male  
 LOCUS BB872020 RIKEN full-length enriched, 16 days neonate male  
 DEFINITION BB872020.1 GI:17118230  
 ACCESSION BB872020  
 VERSION BB872020.1  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 380)

REFERENCE 1 (bases 1 to 380)  
 AUTHORS Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hirakawa, T., Hirose, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Seto, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watanabe, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

TITLE Unpublished  
 JOURNAL Unpublished  
 COMMENT Contact: Yoshinide Hayaishizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

FEATURES  
 source

1.380  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G630034A16"  
 /sex="male"  
 /rname\_type="diencephalon"  
 /dev\_stage="16 days neonate"  
 /clone\_lib="RIKEN full-length enriched, 16 days neonate male diencephalon"

BASE COUNT  
 ORIGIN

Query Match 64.7%; Score 19.4; DB 10; Length 380;  
 Best Local Similarity 79.3%; Pred. No. 2.3e+03;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCTGCGTAGTGACCTCTGCAGG 29  
 DB 12 CCTGCGTAGTGACCTCTGCAGG 40

RESULT 11 406 bp mRNA linear EST 24-DEC-1997  
 AA709478- VV34A05.r1 StrataGene mouse heart (#937316) Mus musculus cDNA clone  
 LOCUS AA709478 IMAGE:1224272 5', mRNA sequence.  
 DEFINITION AA709478  
 ACCESSION AA709478  
 VERSION AA709478.1  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 406)

REFERENCE 1 (bases 1 to 406)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kuwaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellin, B., Kuwaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellin, B., Kuwaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
 The WashU-HMI Mouse EST Project

TITLE Unpublished  
 JOURNAL Unpublished  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:649864  
 Seq primer: -28m3 rev1 ET from Amersham  
 High quality sequence stop: 401.  
 Location/Qualifiers

source

1. 406

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="NIH Swiss"

/db\_xref="taxon:10090"

/clone="IMAGE:1224272"

/seq="pooled"

/tissue\_type="heart"

/dev\_stage="13 day embryos"

/lab\_host="SOLR (kanamycin resistant)"

/clone\_lib="Stratagene mouse heart (#97316)"

/note="Organ: heart; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 65 a 117 c 132 g 92 t

ORIGIN

Query Match 64.7%; Score 19.4; DB 9; Length 406;

Best Local Similarity 79.3%; Pred. No. 2.3e+03;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTCGCTAGTGTGACGACCTCTCCAGG 29

Db 168 CCTGTGAGTGTGAGAGAGCTGCTCAGG 196

RESULT 12

LOCUS BY680656 432 bp mRNA linear EST 16-DEC-2002

DEFINITION BY680656 RIKEN full-length enriched, 12.5 days embryo Rathke's

ACCESSION BY680656

VERSION BY680656.1 GI:27073409

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 432)

AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Okazaki, Y., Furuta, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nishida, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, D.A., Brad, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Giesi, C., Godzik, A., Gough, J., Grimmond, S., Guernicich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kikukawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Atakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kugawa, I., Miyazaki, A., Sakai, K., Shibata, K., Shingawa, A., Yasuniishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp

URL: http://genome-gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Nakamura, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse libraries.

Libraries were provided by Michelle Brinkmeier and Sally Camper (Dept. Human Genetics University of Michigan Medical School 4301 MSB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA ) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome-gsc.riken.go.jp>) for further details.

FEATURES

source

1. 432

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CD-1"

/db\_xref="taxon:10090"

/clone="K920033C15"

/tissue\_type="Rathke's pouches"

/dev\_stage="12.5 days embryo"

/clone\_lib="RIKEN full-length enriched, 12.5 days embryo Rathke's pouches"

BASE COUNT 114 a 100 c 84 g 134 t

ORIGIN

Query Match 64.7%; Score 19.4; DB 13; Length 432;

Best Local Similarity 79.3%; Pred. No. 2.3e+03;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTCGCTAGTGTGACGACCTCTCCAGG 29

Db 118 CCTGCTGTGAGTGTGACGACCTCTCCAGG 146

RESULT 13

LOCUS CB790399 438 bp mRNA linear EST 16-MAY-2003

DEFINITION AMGNUNC:NP14-00017-D1-A W Rat pituitary (10472) Rattus norvegicus

ACCESSION CB790399

VERSION CB790399.1 GI:29878791

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 438)  
 AUTHORS Amgen EST Program.  
 TITLE Amgen Rat EST Program  
 JOURNAL Unpublished  
 COMMENT Contact: Dan Fitzpatrick  
 Amgen, Inc  
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00017 row: d column: 1.  
 Location/Qualifiers  
 1.438  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrp14-00017-d1"  
 /clone\_1lb="W Rat pituitary (10472)"  
 /note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; W Rat pituitary adult female Wistar rat avg insert size 2.1 kb"

BASE COUNT 117 a 93 c 121 g 70 t  
 ORIGIN

Query Match 64.7%; Score 19.4; DB 14; Length 438;  
 Best Local Similarity 79.3%; Pred. No. 2.3e+03;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTACGACCTCTCCGAG 29  
 |||||  
 Db 331 CTCTGAGTCTCGAGGAGCTCTCTCGAG 359

RESULT 14  
 LOCUS BB857961 474 bp mRNA linear EST 26-NOV-2001  
 DEFINITION BB857961.1 RIKEN full-length enriched, B16 F10Y cells Mus musculus  
 CDNA clone G370047P05 5', mRNA sequence.  
 BB857961.1 GI:17099415  
 EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 474)  
 AKimura,T., Arikawa,T., Carinici,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imocani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Komori,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numata,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length CDNA (Akimura,T., et al. 2001)

TITLE Unpublished  
 JOURNAL Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 E-mail: genome-gsc.riken.go.jp/  
 URL: http://genome-gsc.riken.go.jp/  
 Carlinici,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komori,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wael,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Komori,H., Fukunishi,Y., Shibata,K., Itoh,M., Carinici,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (http://genome-gsc.riken.go.jp) for further details.  
 e mouse tissues.  
 Location/Qualifiers  
 1.474  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G370047P05"  
 /clone\_1lb="B16 F10Y cells"  
 /cell\_type="B16 F10Y cells"  
 /clone\_1lb="RIKEN full-length enriched, B16 F10Y cells"

BASE COUNT 73 a 119 c 161 g 121 t  
 ORIGIN

Query Match 64.7%; Score 19.4; DB 10; Length 474;  
 Best Local Similarity 79.3%; Pred. No. 2.3e+03;  
 Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTACGACCTCTCCGAG 29  
 |||||  
 Db 37 CCTGTGAGTGTGAGGAGCTCTCTCGAG 65

RESULT 15  
 LOCUS BF022130 531 bp mRNA linear EST 29-DEC-2000  
 DEFINITION BF022130.1 NCI CGAP Lu30 Mus musculus cDNA clone IMAGE:3662683 5' similar to TR:Q9WTM1.Q9WTM1 MROCT.1, mRNA sequence.  
 BF022130.1 GI:10753462  
 EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 531)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

TITLE Unpublished  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-rcmail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 image.llnl.gov/image/html/resources.shtml  
 MG1:1423451  
 Seq primer: -40RP from Gibco  
 High quality sequence eTop: 436.  
 Location/Qualifiers  
 1.531  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3662683"  
 /tissue\_type="tumor, metastatic to mammary"  
 /lab\_host="DH10B"  
 /clone\_1lb="NCI-CGAP Lu30"  
 /note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; transgenic model WNT-1, expression driven by



MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo  
GT. Library constructed by Life Technologies.  
Investigator providing samples: Gilbert Smith, NIH"  
BASE COUNT 66 a 184 c 169 g 112 t  
ORIGIN

Query Match 64.7%; Score 19.4; DB 10; Length 531;  
Best Local Similarity 79.3%; Pred.No. 2.4e+03;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Oy 1 CCTGCTAGTGTACGACCTCTGCAGG 29  
|||  
Db 409 CCTGTGAGTGGAGAGAGCTGCTGCAGG 437

Search completed: September 16, 2003, 20:39:57  
Job time : 1522.42 secs

BEST AVAILABLE COPY

Wed Sep 17 10:20:10 2003

US-09-594-065-1.rnrm

Schenner L.  
09/15/03  
Seq. IDs 14 w/headers  
Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:28:15 ; Search time 1450.49 Seconds  
(without alignments)  
457.275 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20  
Sequence: 1 tcaccaccgtcagcaccctc 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3363688 seqs, 1658189874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents NA Main:\*

1: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/PCTUS\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US08\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US092\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US092B\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US093B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US093B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US095\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US096\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq.\*  
29: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US098D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US099D\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US099E\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US099F\_COMB.seq.\*

SUMMARIES			
Result No.	Score	Query Match Length	ID
1	20	100.0	US-09-594-065-1
2	20	100.0	US-09-594-065-8
3	20	100.0	US-09-594-065-12
4	20	100.0	US-09-594-065-13

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```
5 20 100.0 109 25 US-09-594-065-14 Sequence 14, Appl
6 20 100.0 481 34 US-09-821-837-9580 Sequence 9580, Ap
7 20 100.0 495 34 US-09-821-837-9581 Sequence 9581, Ap
8 20 100.0 2645 6 US-08-123-456-141 Sequence 141, App
9 20 100.0 2645 18 US-09-297-477A-141 Sequence 141, App
10 20 100.0 2645 43 US-09-994-404-141 Sequence 141, App
11 20 100.0 2713 13 US-08-804-439-13 Sequence 13, Appl
12 20 100.0 2713 19 US-09-301-390-13 Sequence 13, Appl
13 20 100.0 2713 19 US-09-338-328-13 Sequence 13, Appl
14 20 100.0 2275 96 US-60-413-956-11 Sequence 11, Appl
15 20 100.0 2841 6 US-08-123-456-78 Sequence 78, Appl
16 20 100.0 2841 18 US-09-297-477A-78 Sequence 78, Appl
17 20 100.0 2841 43 US-09-994-404-78 Sequence 78, Appl
18 20 100.0 117213 6 US-08-123-456-217 Sequence 217, App
19 20 100.0 117213 18 US-09-297-477A-217 Sequence 217, App
20 20 100.0 117213 43 US-09-994-404-217 Sequence 217, App
21 20 100.0 154746 1 PCT-US01-11372-8 Sequence 8, Appl1
22 20 100.0 154746 2 PCT-US01-11372-8 Sequence 8, Appl1
23 20 100.0 154746 34 US-09-827-688-8 Sequence 8, Appl1
24 20 100.0 34094 42 US-09-963-038A-1 Sequence 1, Appl1
25 20 100.0 34094 47 US-10-199-550-1 Sequence 1, Appl1
26 20 100.0 34094 48 US-10-245-603A-1 Sequence 1, Appl1
27 20 100.0 34094 62 US-60-081-882-1 Sequence 1, Appl1
28 20 100.0 302 83 US-60-288-292-11509 Sequence 11509, A
29 20 100.0 354 28 US-09-654-617-224251 Sequence 224251, A
30 20 100.0 354 30 US-09-684-016-224251 Sequence 224251, A
31 20 100.0 387 34 US-09-804-730-22459 Sequence 22459, A
32 20 100.0 387 72 US-60-189-657-22453 Sequence 22453, A
33 20 100.0 415 34 US-09-804-730-21290 Sequence 21290, A
34 20 100.0 415 72 US-60-189-657-21284 Sequence 21284, A
35 20 100.0 420 44 US-10-020-338-3102 Sequence 3102, Ap
36 20 100.0 428 34 US-09-804-730-20654 Sequence 20654, A
37 20 100.0 428 72 US-60-189-657-20648 Sequence 20648, A
38 20 100.0 471 71 US-60-172-376-1446 Sequence 1446, Ap
39 20 100.0 476 28 US-09-654-617-130864 Sequence 130864, A
40 20 100.0 476 30 US-09-684-016-130864 Sequence 130864, A
41 20 100.0 488 25 US-09-565-309A-1706 Sequence 1706, Ap
42 20 100.0 511 25 US-09-565-309A-44343 Sequence 44343, A
43 20 100.0 523 22 US-09-505-533-38163 Sequence 38163, A
44 20 100.0 523 34 US-09-819-091A-38163 Sequence 38163, A
45 20 100.0 541 71 US-60-172-519-8571 Sequence 8571, Ap
```

## ALIGNMENTS

```
RESULT 1
US-09-594-065-1
; Sequence 1, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: Walkerpach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Forward PCR primer
US-09-594-065-1
```

Query Match 100.0%; Score 20; DB 25; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCACCAACCGTCAGCACCTTC 20

|||||

Db 1 TCACCAACCGTCAGCACCTTC 20

## RESULT 2

```
US-09-594-065-8
; Sequence 8, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: Walkerpach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 a
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 74
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Forward primer for synthesis of IAC
US-09-594-065-8
```

Query Match 100.0%; Score 20; DB 25; Length 74;

Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCACCAACCGTCAGCACCTTC 20

|||||

10 TCACCAACCGTCAGCACCTTC 29

## RESULT 3

```
US-09-594-065-12
; Sequence 12, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: Walkerpach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 a
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 109
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence amplified by PCR reaction for HSV type-1
US-09-594-065-12
```

Query Match 100.0%; Score 20; DB 25; Length 109;

Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCACCAACCGTCAGCACCTTC 20

|||||

1 TCACCAACCGTCAGCACCTTC 20

## RESULT 4

```
US-09-594-065-13
; Sequence 13, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: Walkerpach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 a
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
```

CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 13  
LENGTH: 109  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Sequence amplified by PCR for HSV type-2  
US-09-594-065-13

Query Match 100.0%; Score 20; DB 25; Length 109;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 1 TCACCACCGTCAGCACCTTC 20

RESULT 5  
US-09-594-065-14  
Sequence 14, Application US/09594065  
GENERAL INFORMATION:  
APPLICANT: Walkerpeach, Cindy  
APPLICANT: Dubois, Dwight  
TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 and  
FILE REFERENCE: 25436/1280  
CURRENT APPLICATION NUMBER: US/09/594,065  
CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 14  
LENGTH: 109  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Sequence amplified by PCR for IAC  
US-09-594-065-14

Query Match 100.0%; Score 20; DB 25; Length 109;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 1 TCACCACCGTCAGCACCTTC 20

RESULT 6  
US-09-821-837-9580/C  
Sequence 9580, Application US/09821837  
GENERAL INFORMATION:  
APPLICANT: Gearling, David P.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
FILE REFERENCE: 1600.2076-001  
CURRENT APPLICATION NUMBER: US/09/821,837  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: US 60/193,481  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 9928  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9580  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(481)  
OTHER INFORMATION: n = A,T,C or G

US-09-821-837-9580

Query Match 100.0%; Score 20; DB 34; Length 481;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 381 TCACCACCGTCAGCACCTTC 362

RESULT 7  
US-09-821-837-9581  
Sequence 9581, Application US/09821837  
GENERAL INFORMATION:  
APPLICANT: Gearling, David P.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
FILE REFERENCE: 1600.2076-001  
CURRENT APPLICATION NUMBER: US/09/821,837  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: US 60/193,481  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 9928  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9581  
LENGTH: 495  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(495)  
OTHER INFORMATION: n = A,T,C or G  
US-09-821-837-9581

Query Match 100.0%; Score 20; DB 34; Length 495;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 159 TCACCACCGTCAGCACCTTC 178

RESULT 8  
US-08-123-456-141/C  
Sequence 141, Application US/08123456  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLOO, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/08/123,456  
CURRENT FILING DATE: 1997-03-03  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 141  
LENGTH: 2645  
TYPE: DNA  
ORGANISM: Herpes simplex  
US-08-123-456-141

Query Match 100.0%; Score 20; DB 6; Length 2645;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGACCTTC 20  
DB 814 TCACCAACGTCAGACCTTC 795

RESULT 9  
US-09-297-477A-141/c  
Sequence 141, Application US/09297477A

GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/09/297,477A  
CURRENT FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 141  
LENGTH: 2645  
TYPE: DNA  
ORGANISM: Herpes simplex  
US-09-297-477A-141

Query Match 100.0%; Score 20; DB 18; Length 2645;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGACCTTC 20  
DB 814 TCACCAACGTCAGACCTTC 795

RESULT 10  
US-09-994-404-141/c  
Sequence 141, Application US/09994404  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/09/994,404  
CURRENT FILING DATE: 2001-11-26  
PRIOR APPLICATION NUMBER: US/09/297,477  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 141  
LENGTH: 2645  
TYPE: DNA

ORGANISM: Herpes simplex  
US-09-994-404-141

Query Match 100.0%; Score 20; DB 43; Length 2645;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGACCTTC 20  
DB 814 TCACCAACGTCAGACCTTC 795

RESULT 11  
US-08-804-439-13  
Sequence 13, Application US/08804439

GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARIN L.  
APPLICANT: STRAND, KURT  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RPHV/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSER: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,439  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SCHIEF, J. MICHAEL  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20002.20  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-804-439-13

Query Match 100.0%; Score 20; DB 13; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGACCTTC 20  
DB 1988 TCACCAACGTCAGACCTTC 2007

RESULT 12  
US-09-301-390-13  
Sequence 13, Application US/09301390  
GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARIN L.  
APPLICANT: STRAND, KURT  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RPHV/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 100

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSER: Morrison & Foerster  
;; STREET: 755 Page Mill Road  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304-1018  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/301,390  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/720,229  
;; FILING DATE: 26-SEP-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Schiff, J. Michael  
;; REGISTRATION NUMBER: 40,253  
;; REFERENCE/DOCKET NUMBER: 29938-20002.00  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 813-5600  
;; TELEFAX: (415) 494-0792  
;; TELEX: 706141  
;; INFORMATION FOR SEQ ID NO: 13:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2713 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-09-301-390-13

Query Match 100.0%; Score 20; DB 19; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
Db 1988 TCACCACCGTCAGCACCTTC 2007

RESULT 13  
US-09-338-326-13  
;; Sequence 13, Application US/09338326  
;; GENERAL INFORMATION:  
;; APPLICANT: Rose, Timothy M.  
;; APPLICANT: Bosch, Marnix L.  
;; APPLICANT: Strand, Kurt  
;; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RRV/KSHV  
;; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
;; NUMBER OF SEQUENCES: 100  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSER: Morrison & Foerster  
;; STREET: 755 Page Mill Road  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304-1018  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/338,326  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/720,229  
;; FILING DATE: 26-SEP-1996

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Schiff, J. Michael  
;; REGISTRATION NUMBER: 40,253  
;; REFERENCE/DOCKET NUMBER: 29938-20002.00  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 813-5600  
;; TELEFAX: (415) 494-0792  
;; TELEX: 706141  
;; INFORMATION FOR SEQ ID NO: 13:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2713 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-09-338-326-13

Query Match 100.0%; Score 20; DB 19; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
Db 1988 TCACCACCGTCAGCACCTTC 2007

RESULT 14  
US-60-412-956-11  
;; Sequence 11, Application US/60412956  
;; GENERAL INFORMATION:  
;; APPLICANT: SYKES, KATHRYN F.  
;; APPLICANT: STEWKE-HALE, KATHERINE  
;; APPLICANT: JOHNSTON, STEPHEN ALBERT  
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING  
;; TITLE OF INVENTION: HERPESVIRUS FAMILY  
;; FILE REFERENCE: MCRO:002USP1  
;; CURRENT APPLICATION NUMBER: US/60/412,956  
;; CURRENT FILING DATE: 2002-09-23  
;; NUMBER OF SEQ ID NOS: 24  
;; SOFTWARE: Patent In Ver. 2.1  
;; SEQ ID NO 11  
;; LENGTH: 2715  
;; TYPE: DNA  
;; ORGANISM: HERPES VIRUS, TYPE 1  
;; US-60-412-956-11

Query Match 100.0%; Score 20; DB 96; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
Db 1991 TCACCACCGTCAGCACCTTC 2010

RESULT 15  
US-08-123-456-78  
;; Sequence 78, Application US/08123456  
;; GENERAL INFORMATION:  
;; APPLICANT: ESSER, KLAUS M.  
;; APPLICANT: CHAN, JOHN Y.  
;; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
;; APPLICANT: DELVECCHIO, ALFRED MICHAEL  
;; APPLICANT: DILLON, SUSAN B.  
;; APPLICANT: LEARY, JEFFREY JOSEPH  
;; APPLICANT: SUTTON, DAVID  
;; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
;; TITLE OF INVENTION: SIMPLEX VIRUS TYPE-2  
;; FILE REFERENCE: P50583  
;; CURRENT APPLICATION NUMBER: US/08/123,456  
;; CURRENT FILING DATE: 1997-03-03  
;; PRIOR APPLICATION NUMBER: US 60/030,279  
;; PRIOR FILING DATE: 1999-11-04

/ PRIOR APPLICATION NUMBER: US 60/049,018  
/ PRIOR FILING DATE: 1997-06-09  
/ NUMBER OF SEQ ID NOS: 303  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 78  
/ LENGTH: 2841  
/ TYPE: DNA  
/ ORGANISM: Herpes simplex  
/ FEATURE:  
/ NAME/KEY: unknown  
/ LOCATION: (2591)(2825)(2833)  
/ OTHER INFORMATION:  
US-08-123-456-78

Query Match 100.0%; Score 20; DB 6; Length 2841;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACCAACCGTCACGACCTTC 20  
Db 2012 TCACCAACCGTCACGACCTTC 2031

Search completed: September 16, 2003, 22:44:43  
Job time : 1452.49 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 18:37:15 ; Search time 31.4851 Seconds  
(without alignments)  
535.803 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20  
Sequence: 1 tcaccacgcgcagcacttc 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 606776 seqs, 421745864 residues

Total number of hits satisfying chosen parameters: 1213552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New.\*  
1: /cgn2\_6/ptodata/2/pna/PCr\_NEW\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	527	6 US-10-425-114A-27295	Sequence 27295, A
2	16.8	84.0	2086	6 US-10-425-114A-21287	Sequence 21287, A
3	16.8	84.0	7838	5 US-09-654-936A-147	Sequence 147, App
4	16.8	84.0	91672	7 US-60-487-610-19444	Sequence 19444, A
5	16.8	84.0	2731748	6 US-10-297-465B-1	Sequence 1, Appli
6	16.4	82.0	25086	7 US-60-487-610-18609	Sequence 19609, A
7	16.4	82.0	25474	7 US-60-495-114-16832	Sequence 16832, A
8	16.4	82.0	207149	7 US-60-487-610-19377	Sequence 19377, A
9	16	80.0	614	6 US-10-425-114A-19377	Sequence 19377, A
10	15.8	79.0	179	1 PCr-US03-18714-26075	Sequence 26075, A
11	15.8	79.0	630	1 PCr-US02-37235-115	Sequence 115, App
12	15.8	79.0	749	6 US-10-425-114A-7593	Sequence 7593, App
13	15.8	79.0	772	6 US-10-425-114A-14861	Sequence 14861, A
14	15.8	79.0	784	6 US-10-425-114A-26570	Sequence 26570, A
15	15.8	79.0	919	6 US-10-425-114A-24003	Sequence 24003, A
16	15.8	79.0	2264	6 US-10-425-114A-19130	Sequence 19130, A
17	15.8	79.0	2533	6 US-10-425-114A-35600	Sequence 35600, A
18	15.8	79.0	2649	6 US-10-425-114A-24312	Sequence 24312, A
19	15.8	79.0	4000	5 US-09-976-858-281	Sequence 281, App
20	15.8	79.0	6014	7 US-60-490-890-2216	Sequence 2216, App
21	15.8	79.0	11740	7 US-60-490-890-706	Sequence 706, App
22	15.8	79.0	142005	7 US-60-487-610-19647	Sequence 19647, A
23	15.8	79.0	200820	7 US-60-495-114-16936	Sequence 16936, A
24	15.8	79.0	4813087	5 US-09-947-914-75	Sequence 75, Appl
25	15.4	77.0	1271	6 US-10-425-114A-10829	Sequence 10829, A
26	15.4	77.0	1371	6 US-10-425-114A-11773	Sequence 11773, A

27	15.4	77.0	1972	6 US-10-425-114A-20032	Sequence 20032, A
28	15.4	77.0	2023	6 US-10-425-114A-10611	Sequence 10611, A
29	15.4	77.0	76798	7 US-60-490-890-1467	Sequence 1467, App
30	15.4	77.0	201007	7 US-60-487-610-19435	Sequence 19435, A
31	15.4	77.0	255673	7 US-60-495-114-15366	Sequence 15366, A
32	15.4	77.0	255673	7 US-60-495-135-3613	Sequence 3613, App
33	15.4	77.0	513030	7 US-60-487-610-19737	Sequence 19737, A
34	15.4	77.0	2731748	6 US-10-297-465B-1	Sequence 1, Appli
35	15.2	76.0	201	7 US-60-487-610-40113	Sequence 40113, A
36	15.2	76.0	201	7 US-60-487-610-40114	Sequence 40114, A
37	15.2	76.0	201	7 US-60-487-610-40115	Sequence 40115, A
38	15.2	76.0	201	7 US-60-487-610-40116	Sequence 40116, A
39	15.2	76.0	201	7 US-60-487-610-58562	Sequence 58562, A
40	15.2	76.0	201	7 US-60-487-610-69696	Sequence 69696, A
41	15.2	76.0	201	7 US-60-487-610-69700	Sequence 69700, A
42	15.2	76.0	201	7 US-60-487-610-69730	Sequence 69730, A
43	15.2	76.0	201	7 US-60-487-610-69747	Sequence 69747, A
44	15.2	76.0	201	7 US-60-487-610-69753	Sequence 69753, A
45	15.2	76.0	201	7 US-60-487-610-70283	Sequence 70283, A

## ALIGNMENTS

```
RESULT 1
US-10-425-114A-27295/c
; Sequence 27295, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 27295
; LENGTH: 527
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURES:
; OTHER INFORMATION: Clone ID: LIB4729-056-B3_FLI
US-10-425-114A-27295

Query Match      84.0% Score 16.8; DB 6; Length 527;
Best Local Similarity 90.0% Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCACCACGCTGACGACCTTC 20
DB      264 TCACCACGCTGACGACCTTC 245

RESULT 2
US-10-425-114A-21287/c
; Sequence 21287, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425, 114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
```

SEQ ID NO 21287  
LENGTH: 2086  
TYPE: DNA  
ORGANISM: Zee mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3279-218-A5\_FLI  
US-10-425-114A-21287

Query Match 84.0%; Score 16.8; DB 6; Length 2086;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCACCACTGTCAGCACCCTC 20  
Db 620 TCACCACTGTCAGCACCCTC 601

RESULT 3  
US-09-654-936A-147/c  
Sequence 147, Application US/09654936A  
GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhou, Ping  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyun  
APPLICANT: Zhao, Qing A.  
APPLICANT: Xue, Aidong J.  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Dmanac, Radoje T.  
TITLE OF INVENTION: Novel Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 787CIP2C  
CURRENT APPLICATION NUMBER: US/09/654,936A  
CURRENT FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 09/560,875  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 09/496,914  
PRIOR FILING DATE: 2000-02-03  
NUMBER OF SEQ ID NOS: 164  
SOFTWARE: pc\_fl\_gene Version 1.0  
SEQ ID NO 147  
LENGTH: 7838  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (673)..(2616)  
US-09-654-936A-147

Query Match 84.0%; Score 16.8; DB 5; Length 7838;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCACCACTGTCAGCACCCTC 20  
Db 6499 TCACCACTGTCAGCACCCTC 6480

RESULT 4  
US-60-487-610-19444  
Sequence 19444, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: HUANG, Hongjin  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,  
TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17

NUMBER OF SEQ ID NOS: 97101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19444  
LENGTH: 91672  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(91672)  
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1  
US-60-487-610-19444

Query Match 84.0%; Score 16.8; DB 7; Length 91672;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCACCACTGTCAGCACCCTC 20  
Db 19866 TCACCACTGTCAGCACCCTC 19885

RESULT 5  
US-10-297-465B-1/c  
Sequence 1, Application US/10297465B  
GENERAL INFORMATION:  
APPLICANT: Simpson, Andrew  
APPLICANT: Reinach, Fernando  
APPLICANT: Setubal, Joao  
APPLICANT: Mediane, Joao  
APPLICANT: Arruda, Paulo  
TITLE OF INVENTION: Isolated Genome of *Xylella fastidiosa* and Uses Thereof  
FILE REFERENCE: FAPESP 202 US (10213376)  
CURRENT APPLICATION NUMBER: US/10/297,465B  
CURRENT FILING DATE: 2003-08-25  
PRIOR APPLICATION NUMBER: PCT/IB01/01618  
PRIOR FILING DATE: 2001-06-07  
PRIOR APPLICATION NUMBER: 60/209,906  
PRIOR FILING DATE: 2001-06-17  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 2731748  
TYPE: DNA  
ORGANISM: *Xylella fastidiosa*  
US-10-297-465B-1

Query Match 84.0%; Score 16.8; DB 6; Length 2731748;  
Best Local Similarity 90.0%; Pred. No. 99;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCACCACTGTCAGCACCCTC 20  
Db 723037 TCACCACTGTCAGCACCCTC 723018

RESULT 6  
US-60-487-610-19609  
Sequence 19609, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: HUANG, Hongjin  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,  
TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17  
NUMBER OF SEQ ID NOS: 97101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19609  
LENGTH: 25086  
TYPE: DNA  
ORGANISM: Homo sapiens

US-60-487-610-19609

Query Match 82.0%; Score 16.4; DB 7; Length 25086;  
Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CACCACCGTCGACCTT 19  
|||||  
DB 203 CACCACCATCAGACCTT 220

RESULT 7

US-60-495-114-16832  
; Sequence 16832, Application US/60495114  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO01480  
; CURRENT APPLICATION NUMBER: US/60/495,114  
; CURRENT FILING DATE: 2003-08-15  
; NUMBER OF SEQ ID NOS: 91238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16832  
; LENGTH: 25474  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(25474)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-60-495-114-16832

Query Match 82.0%; Score 16.4; DB 7; Length 25474;  
Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CACCACCGTCGACCTT 19  
|||||  
DB 298 CACCACCATCAGACCTT 315

RESULT 8

US-60-487-610-19377/C  
; Sequence 19377, Application US/60487610  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INJECTED SUBJECTS,  
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01469  
; CURRENT APPLICATION NUMBER: US/60/487,610  
; CURRENT FILING DATE: 2003-07-17  
; NUMBER OF SEQ ID NOS: 97101  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19377  
; LENGTH: 207149  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(207149)  
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
US-60-487-610-19377

Query Match 82.0%; Score 16.4; DB 7; Length 207149;  
Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCGACCT 18  
|||||

DB 86537 TCACCACTGTCGACCT 86520

RESULT 9

US-10-425-114A-19377/C  
; Sequence 19377, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabeska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 19377  
; LENGTH: 614  
; TYPE: DNA  
; ORGANISM: Gossypium hirsutum  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3146-010-H6\_FLI  
US-10-425-114A-19377

Query Match 80.0%; Score 16; DB 6; Length 614;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CACCACCGTCGACCTT 17  
|||||  
DB 213 CACCACCGTCGACCTT 198

RESULT 10

PCT-US03-18714-26075/C  
; Sequence 26075, Application PC/TUS0318714  
; GENERAL INFORMATION:  
; APPLICANT: Regulome Corporation  
; TITLE OF INVENTION: Functional Sites  
; FILE REFERENCE: 11207-029-228  
; CURRENT APPLICATION NUMBER: PCT/US03/18714  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/387,887  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/387,910  
; PRIOR FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 51999  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 26075  
; LENGTH: 179  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
PCT-US03-18714-26075

Query Match 79.0%; Score 15.8; DB 1; Length 179;  
Best Local Similarity 89.5%; Pred. No. 3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CACCACCGTCGACCTT 20  
|||||  
DB 29 CTCACCGTCGACCTT 11

RESULT 11

PCT-US02-37235-115  
; Sequence 115, Application PC/TUS0237235  
; GENERAL INFORMATION:  
; APPLICANT: Handfield, Martin  
; APPLICANT: Hillman, Jeffrey  
; APPLICANT: Proguiske-Fox, Ann

```
/ TITLE OF INVENTION: Identification of Actinobacillus actinomycetemcomitans Antigens
/ FILE REFERENCE: MBH901-662A
/ CURRENT APPLICATION NUMBER: PCT/US02/37235
/ NUMBER OF SEQ ID NOS: 234
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 115
/ LENGTH: 630
/ TYPE: DNA
/ ORGANISM: Actinobacillus actinomycetemcomitans
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(630)
PCT-US02-37235-115

Query Match          79.0%; Score 15.8; DB 1; Length 630;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TCACGACCGTCAGCACCTT 19
DB      386 TCGCCACCGTTACGACCTT 404

RESULT 12
US-10-425-114A-7593/C
/ Sequence 7593, Application US/10425114A
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(5313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114A
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 7593
/ LENGTH: 749
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: 700674452_FLI
US-10-425-114A-7593

Query Match          79.0%; Score 15.8; DB 6; Length 749;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CACGACCGTCAGCACCTTC 20
DB      584 CACGACCATCAGCACCTTC 566

RESULT 13
US-10-425-114A-14861/C
/ Sequence 14861, Application US/10425114A
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(5313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114A
/ CURRENT FILING DATE: 2003-04-28
```

```
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 14861
/ LENGTH: 772
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB3029-011-F8_FLI
US-10-425-114A-14861

Query Match          79.0%; Score 15.8; DB 6; Length 772;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CACGACCGTCAGCACCTTC 20
DB      608 CACGACCATCAGCACCTTC 590

RESULT 14
US-10-425-114A-26570/C
/ Sequence 26570, Application US/10425114A
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(5313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114A
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 26570
/ LENGTH: 784
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB4315-040-D4_FLI
US-10-425-114A-26570

Query Match          79.0%; Score 15.8; DB 6; Length 784;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CACGACCGTCAGCACCTTC 20
DB      147 CACGACCTTCAGCGCCTTC 129

RESULT 15
US-10-425-114A-24003/C
/ Sequence 24003, Application US/10425114A
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(5313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114A
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 24003
/ LENGTH: 919
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB3606-055-G3_FLI
```

US-10-425-114A-24003

Query Match 79.0%; Score 15.8; DB 6; Length 919;

Best Local Similarity 89.5%; Pred. No. 3e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TCACGACCGTCAGACCTT 19

Db 634 TCACCTCCTTCAGACCTT 616

Search completed: September 16, 2003, 22:47:51  
Job time : 41.4851 secs

**BEST AVAILABLE COPY**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:28:15 ; Search time 1305.45 Seconds

(without alignments)  
457.275 Million cell updates/sec

Title: US-09-594-065-2

Perfect score: 18  
Sequence: 1 agcaggccgcctgccttg 18

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 33363688 seqs, 1658189874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

```

Pending_Patents_NA_Main:*
1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US092A_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US092B_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US093A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US098D_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*

```

Result				SUMMARIES			
No.	Score	Match	Query length	ID	Description		
1	18	100.0	18	25	US-09-594-065-2	Sequence 2, Appl	
2	18	100.0	73	25	US-09-594-065-9	Sequence 9, Appl	
3	18	100.0	109	25	US-09-594-065-12	Sequence 12, Appl	
4	18	100.0	109	25	US-09-594-065-13	Sequence 13, Appl	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```

c 5 18 100.0 109 25 US-09-594-065-14 Sequence 14, Appl
6 18 100.0 481 34 US-09-821-837-9580 Sequence 9580, Ap
c 7 18 100.0 495 34 US-09-821-837-9581 Sequence 9581, Ap
8 18 100.0 2645 6 US-08-123-456-141 Sequence 141, App
9 18 100.0 2645 18 US-09-297-477A-141 Sequence 141, App
10 18 100.0 2645 43 US-09-994-404-141 Sequence 141, App
c 11 18 100.0 2713 13 US-08-804-439-13 Sequence 13, Appl
c 12 18 100.0 2713 13 US-09-301-390-13 Sequence 13, Appl
c 13 18 100.0 2713 19 US-09-338-326-13 Sequence 13, Appl
c 14 18 100.0 2715 96 US-60-412-956-11 Sequence 11, Appl
c 15 18 100.0 2841 6 US-08-123-456-78 Sequence 78, Appl
c 16 18 100.0 2841 18 US-09-297-477A-78 Sequence 78, Appl
c 17 18 100.0 2841 43 US-09-994-404-78 Sequence 78, Appl
c 18 100.0 2943 10 US-08-541-878-7 Sequence 7, Appl
19 18 100.0 117213 6 US-08-123-456-217 Sequence 217, App
20 18 100.0 117213 18 US-09-297-477A-217 Sequence 217, App
21 18 100.0 117213 43 US-09-994-404-217 Sequence 217, App
22 18 100.0 154746 1 PCT-US01-11372-8 Sequence 8, Appl
23 18 100.0 154746 2 PCT-US01-11372-8 Sequence 8, Appl
24 18 100.0 154746 34 US-09-827-688-8 Sequence 1201, Ap
c 25 17 94.4 855 51 US-10-366-683-1201 Sequence 1201, Ap
26 17 94.4 855 52 US-10-419-128-1201 Sequence 7812, Ap
c 27 17 94.4 1431 1 PCT-US02-03987-7812 Sequence 7812, Ap
c 28 17 94.4 1431 34 US-09-815-242-7812 Sequence 7812, Ap
c 29 17 94.4 1431 45 US-10-072-851-7812 Sequence 30286, A
c 30 17 94.4 1431 49 US-10-282-122A-30286 Sequence 1243, Ap
c 31 17 94.4 1491 51 US-10-366-683-1243 Sequence 1243, Ap
c 32 17 94.4 1551 52 US-10-419-128-1243 Sequence 1352, Ap
33 17 94.4 1551 51 US-10-366-683-1352 Sequence 1352, Ap
34 17 94.4 1551 52 US-10-419-128-1352 Sequence 8, Appl
c 35 17 94.4 6244 9 US-08-076-327-8 Sequence 8, Appl
c 36 17 94.4 6244 18 US-09-281-674-8 Sequence 8, Appl
c 37 17 94.4 6244 33 US-09-777-317-8 Sequence 8, Appl
c 38 17 94.4 6244 37 US-09-892-227-8 Sequence 601, App
c 39 17 94.4 6244 62 US-60-082-302-601 Sequence 35472, A
c 40 17 94.4 10351 19 1 PCT-US02-25943-35472 Sequence 35472, A
c 41 16.4 91.1 19 48 US-10-227-565-35472 Sequence 35472, A
c 42 16.4 91.1 19 51 US-10-367-832A-35472 Sequence 35472, A
c 43 16.4 91.1 201 100 US-60-452-680-71403 Sequence 71403, A
c 44 16.4 91.1 201 100 US-60-452-680-71403 Sequence 71404, A
c 45 16.4 91.1 201 100 US-60-452-680-71404

```

## ALIGNMENTS

```

RESULT 1
US-09-594-065-2
; Sequence 2, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: WalkerPeach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and Kits for Herpes Simplex Virus Type 1 an
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 18
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Reverse PCR primer
US-09-594-065-2

```

```

Query Match 100.0%; Score 18; DB 25; Length 18;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AGCAGCCCGCTGTCTTG 18
|||||

```

```

Db 1 AGCAGCCCGCTGTCTTG 18
RESULT 2
US-09-594-065-9
; Sequence 9, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: WalkerPeach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and Kits for Herpes Simplex Virus Type 1 a
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 73
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Reverse primer for synthesis of IAC
US-09-594-065-9

```

```

Query Match 100.0%; Score 18; DB 25; Length 73;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AGCAGCCCGCTGTCTTG 18
|||||
10 AGCAGCCCGCTGTCTTG 27

```

```

RESULT 3
US-09-594-065-12/C
; Sequence 12, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: WalkerPeach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and Kits for Herpes Simplex Virus Type 1 an
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 109
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence amplified by PCR reaction for HSV type-1
US-09-594-065-12

```

```

Query Match 100.0%; Score 18; DB 25; Length 109;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 AGCAGCCCGCTGTCTTG 18
|||||
Db 109 AGCAGCCCGCTGTCTTG 92

```

```

RESULT 4
US-09-594-065-13/C
; Sequence 13, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: WalkerPeach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and Kits for Herpes Simplex Virus Type 1 an
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065

```



CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 13  
LENGTH: 109  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Sequence amplified by PCR for HSV type-2  
US-09-594-065-13

Query Match 100.0%; Score 18; DB 25; Length 109;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGGCCGCTGTCTTG 18  
Db 109 AGCAGGCCGCTGTCTTG 92

RESULT 5  
US-09-594-065-14/c  
Sequence 14, Application US/09594065  
GENERAL INFORMATION:  
APPLICANT: WalkerPeach, Cindy  
TITLE OF INVENTION: Compositions, methods and Kits for Herpes Simplex Virus Type 1 an  
FILE REFERENCE: 25436/1280  
CURRENT APPLICATION NUMBER: US/09/594,065  
CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 14  
LENGTH: 109  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Sequence amplified by PCR for IAC  
US-09-594-065-14

Query Match 100.0%; Score 18; DB 25; Length 109;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGGCCGCTGTCTTG 18  
Db 109 AGCAGGCCGCTGTCTTG 92

RESULT 6  
US-09-821-837-9580  
Sequence 9580, Application US/09821837  
GENERAL INFORMATION:  
APPLICANT: Geating, David P.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USBS  
FILE REFERENCE: 1600.2076-001  
CURRENT APPLICATION NUMBER: US/09/821,837  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: US 60/193,481  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 9928  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9580  
LENGTH: 481  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(481)  
OTHER INFORMATION: n = A,T,C or G

US-09-821-837-9580

Query Match 100.0%; Score 18; DB 34; Length 481;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGGCCGCTGTCTTG 18  
Db 273 AGCAGGCCGCTGTCTTG 290

RESULT 7  
US-09-821-837-9581/c  
Sequence 9581, Application US/09821837  
GENERAL INFORMATION:  
APPLICANT: Geating, David P.  
APPLICANT: Holtzman, Douglas A.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USBS  
FILE REFERENCE: 1600.2076-001  
CURRENT APPLICATION NUMBER: US/09/821,837  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: US 60/193,481  
PRIOR FILING DATE: 2000-03-29  
NUMBER OF SEQ ID NOS: 9928  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9581  
LENGTH: 499  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(495)  
OTHER INFORMATION: n = A,T,C or G  
US-09-821-837-9581

Query Match 100.0%; Score 18; DB 34; Length 495;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGGCCGCTGTCTTG 18  
Db 267 AGCAGGCCGCTGTCTTG 250

RESULT 8  
US-08-123-456-141  
Sequence 141, Application US/08123456  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/08/123,456  
CURRENT FILING DATE: 1997-03-03  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 141  
LENGTH: 2645  
TYPE: DNA  
ORGANISM: Herpes simplex  
US-08-123-456-141

Query Match 100.0%; Score 18; DB 6; Length 2645;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCCGCTGTCTTG 18  
|||||  
DB 706 AGCAGCGCCGCTGTCTTG 723

## RESULT 9

US-09-297-477A-141  
Sequence 141, Application US/09297477A  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DILON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
TITLE OF INVENTION: SIMPLEX VIRUS TYPE-2  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/09/297,477A  
CURRENT FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 141  
LENGTH: 2645  
TYPE: DNA  
ORGANISM: Herpes simplex  
US-09-297-477A-141

Query Match 100.0%; Score 18; DB 18; Length 2645;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCCGCTGTCTTG 18  
|||||  
DB 706 AGCAGCGCCGCTGTCTTG 723

## RESULT 10

US-09-994-404-141  
Sequence 141, Application US/09994404  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
TITLE OF INVENTION: SIMPLEX VIRUS TYPE-2  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/09/994,404  
CURRENT FILING DATE: 2001-11-26  
PRIOR APPLICATION NUMBER: 09/297,477  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 141  
LENGTH: 2645  
TYPE: DNA

ORGANISM: Herpes simplex  
US-09-994-404-141

Query Match 100.0%; Score 18; DB 43; Length 2645;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCCGCTGTCTTG 18  
|||||  
DB 706 AGCAGCGCCGCTGTCTTG 723

## RESULT 11

US-08-804-439-13/C  
Sequence 13, Application US/08804439  
GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARIN L.  
APPLICANT: STRAND, KURT  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RPHV/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSER: MORRISON & FOETELER  
STREET: 755 PAGE MILL ROAD  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,439  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: SCHIEF, J. MICHAEL  
REGISTRATION NUMBER: 40,253  
REFERENCE/DOCKET NUMBER: 29938-20002.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-804-439-13

Query Match 100.0%; Score 18; DB 13; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCCGCTGTCTTG 18  
|||||  
DB 2096 AGCAGCGCCGCTGTCTTG 2079

## RESULT 12

US-09-301-390-13/C  
Sequence 13, Application US/09301390  
GENERAL INFORMATION:  
APPLICANT: ROSE, TIMOTHY M.  
APPLICANT: BOSCH, MARIN L.  
APPLICANT: STRAND, KURT  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RPHV/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 100

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Morrison & Foerster  
;; STREET: 755 Page Mill Road  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304-1018  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/301,390  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/720,229  
;; FILING DATE: 26-SEP-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Schief, J. Michael  
;; REGISTRATION NUMBER: 40,253  
;; REFERENCE/DOCKET NUMBER: 29938-20002.00  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 813-5600  
;; TELEFAX: (415) 494-0792  
;; TELERX: 706141  
;; INFORMATION FOR SEQ ID NO: 13:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2713 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-09-301-390-13

Query Match 100.0%; Score 18; DB 19; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGCTCTTG 18  
DB 2096 AGCAGCGCGCTGCTCTTG 2079

RESULT 13  
US-09-338-326-13/C  
;; Sequence 13, Application US/09338326  
;; GENERAL INFORMATION:  
;; APPLICANT: Rose, Timothy M.  
;; APPLICANT: Bosch, Marnix L.  
;; APPLICANT: Strand, Kurt  
;; TITLE OF INVENTION: GLYCOPROTEIN B OF THE PHV/KSHV  
;; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
;; NUMBER OF SEQUENCES: 100  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Morrison & Foerster  
;; STREET: 755 Page Mill Road  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94304-1018  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/338,326  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/720,229  
;; FILING DATE: 26-SEP-1996

;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Schief, J. Michael  
;; REGISTRATION NUMBER: 40,253  
;; REFERENCE/DOCKET NUMBER: 29938-20002.00  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 813-5600  
;; TELEFAX: (415) 494-0792  
;; TELERX: 706141  
;; INFORMATION FOR SEQ ID NO: 13:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2713 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-09-338-326-13

Query Match 100.0%; Score 18; DB 19; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGCTCTTG 18  
DB 2096 AGCAGCGCGCTGCTCTTG 2079

RESULT 14  
US-60-412-956-11/C  
;; Sequence 11, Application US/60412956  
;; GENERAL INFORMATION:  
;; APPLICANT: SYKES, KATHRYN F.  
;; APPLICANT: STEWKE-HALE, KATHERINE  
;; APPLICANT: JOHNSTON, STEPHEN ALBERT  
;; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING  
;; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE  
;; TITLE OF INVENTION: HERPESVIRUS FAMILY  
;; FILE REFERENCE: MICRO:002USP1  
;; CURRENT FILING DATE: 2002-09-23  
;; NUMBER OF SEQ ID NOS: 24  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 11  
;; LENGTH: 2715  
;; TYPE: DNA  
;; ORGANISM: HERPES VIRUS, TYPE 1  
;; US-60-412-956-11

Query Match 100.0%; Score 18; DB 96; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGCTCTTG 18  
DB 2099 AGCAGCGCGCTGCTCTTG 2082

RESULT 15  
US-08-123-456-78/C  
;; Sequence 78, Application US/08123456  
;; GENERAL INFORMATION:  
;; APPLICANT: ESSER, KLAUS M.  
;; APPLICANT: CHAN, JOHN Y.  
;; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
;; APPLICANT: DELVECCHIO, ALFRED MICHAEL  
;; APPLICANT: DILLON, SUSAN B.  
;; APPLICANT: LEAHY, JEFFREY JOSEPH  
;; APPLICANT: SUTTON, DAVID  
;; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
;; TITLE OF INVENTION: SIMPLEX VIRUS TYPE-2  
;; FILE REFERENCE: P50583  
;; CURRENT APPLICATION NUMBER: US/08/123,456  
;; CURRENT FILING DATE: 1997-03-03  
;; PRIOR APPLICATION NUMBER: US 60/030,279  
;; PRIOR FILING DATE: 1999-11-04

;; PRIOR APPLICATION NUMBER: US 60/049,018  
;; PRIOR FILING DATE: 1997-06-09  
;; NUMBER OF SEQ ID NOS: 303  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 78  
;; LENGTH: 2841  
;; TYPE: DNA  
;; ORGANISM: Herpes simplex  
;; FEATURE:  
;; NAME/KEY: unknown  
;; LOCATION: (2591)(2825)(2833)  
;; OTHER INFORMATION:  
US-08-123-456-78

Query Match 100.0%; Score 18; DB 6; Length 2841;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGGCCGCTGCTCTTG 18  
|||  
Db 2120 AGCAGGCCGCTGCTCTTG 2103

Search completed: September 16, 2003, 22:44:45  
Job time : 1307.45 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:37:15 ; Search time 28.3366 Seconds  
(without alignments)  
535.803 Million cell updates/sec

Title: US-09-594-065-2

Perfect score: 18  
Sequence: 1 agcagggcgcgtccttg 18

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 606776 seqs, 421745864 residues

Total number of hits satisfying chosen parameters: 1213552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*  
1: /cgn2\_6/pdata/2/pna/PCT\_NEW\_COMB.seq:\*  
2: /cgn2\_6/pdata/2/pna/US06\_NEW\_COMB.seq:\*  
3: /cgn2\_6/pdata/2/pna/US07\_NEW\_COMB.seq:\*  
4: /cgn2\_6/pdata/2/pna/US08\_NEW\_COMB.seq:\*  
5: /cgn2\_6/pdata/2/pna/US09\_NEW\_COMB.seq:\*  
6: /cgn2\_6/pdata/2/pna/US10\_NEW\_COMB.seq:\*  
7: /cgn2\_6/pdata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.4	91.1	1924	6	US-10-651-237-4 Sequence 4, Appl 1
2	16.4	91.1	1924	6	US-10-651-237-7 Sequence 7, Appl 1
3	15.4	85.6	201	7	US-60-495-114-38730 Sequence 38730, A
4	15.4	85.6	201	7	US-60-495-114-89612 Sequence 89612, A
5	15.4	85.6	201	7	US-60-495-114-89619 Sequence 89619, A
6	15.4	85.6	2441	6	US-10-425-114A-26835 Sequence 26835, A
7	15.4	85.6	2463	6	US-10-425-114A-26959 Sequence 26959, A
8	15.4	85.6	2470	7	US-60-490-890-1861 Sequence 1861, Ap
9	15.4	85.6	2487	6	US-10-425-114A-26285 Sequence 26285, A
10	15.4	85.6	2799	5	US-09-935-368A-2 Sequence 2, Appl 1
11	15.4	85.6	26433	7	US-60-495-114-16387 Sequence 16387, A
12	15.4	85.6	45664	7	US-60-495-114-16957 Sequence 16957, A
13	15.4	85.6	384485	7	US-60-495-114-16382 Sequence 16382, A
14	15.4	83.3	1529	6	US-10-425-114A-27906 Sequence 27906, A
15	14.8	82.2	201	7	US-60-495-114-15241 Sequence 15241, A
16	14.8	82.2	201	7	US-60-495-114-54280 Sequence 54280, A
17	14.8	82.2	201	7	US-60-495-114-54313 Sequence 54313, A
18	14.8	82.2	201	7	US-60-495-114-68094 Sequence 68094, A
19	14.8	82.2	201	7	US-60-495-114-82382 Sequence 82382, A
20	14.8	82.2	201	7	US-60-495-114-82387 Sequence 82387, A
21	14.8	82.2	318	1	PCT-US03-11188-38 Sequence 38, Appl 1
22	14.8	82.2	827	6	US-10-653-047-1247 Sequence 1247, Ap
23	14.8	82.2	924	6	US-10-425-114A-1273 Sequence 1273, Ap
24	14.8	82.2	1066	6	US-10-425-114A-33268 Sequence 33268, A
25	14.8	82.2	1135	6	US-10-425-114A-23292 Sequence 23292, A
26	14.8	82.2	1331	6	US-10-425-114A-27120 Sequence 27120, A

27	14.8	82.2	1373	6	US-10-425-114A-35800 Sequence 35800, A
28	14.8	82.2	1690	6	US-10-425-114A-33240 Sequence 33240, A
29	14.8	82.2	1759	7	US-60-495-114-1078 Sequence 1078, Ap
30	14.8	82.2	1806	6	US-10-425-114A-1679 Sequence 1679, Ap
31	14.8	82.2	1908	6	US-10-425-114A-21804 Sequence 21804, A
32	14.8	82.2	2054	6	US-10-425-114A-28366 Sequence 28366, A
33	14.8	82.2	2492	6	US-10-425-114A-31551 Sequence 31551, A
34	14.8	82.2	2520	6	US-60-482-992-5 Sequence 5, Appl 1
35	14.8	82.2	4212	6	US-10-425-114A-26843 Sequence 26843, A
36	14.8	82.2	22806	7	US-60-495-114-16659 Sequence 16659, A
37	14.8	82.2	39902	7	US-60-487-610-20061 Sequence 20061, A
38	14.8	82.2	53795	1	PCT-US02-38582-124 Sequence 124, App
39	14.8	82.2	64607	7	US-60-495-114-16541 Sequence 16541, A
40	14.8	82.2	84010	7	US-60-487-610-19295 Sequence 19295, A
41	14.8	82.2	111982	7	US-60-487-610-19846 Sequence 19846, A
42	14.8	82.2	144434	7	US-60-495-114-16451 Sequence 16451, A
43	14.4	80.0	201	7	US-60-487-610-86638 Sequence 86638, A
44	14.4	80.0	423	6	US-10-084-846A-107 Sequence 107, App
45	14.4	80.0	1014	1	PCT-US03-23249-113 Sequence 113, App

## ALIGNMENTS

RESULT 1  
US-10-651-237-4/c  
; Sequence 4, Application US/10651237  
; GENERAL INFORMATION:  
; APPLICANT: Ortho-Clinical Diagnostics, Inc.  
; APPLICANT: Wang, Yixin  
; TITLE OF INVENTION: Colorectal Cancer Prognostics  
; FILE REFERENCE: ADS-5003 US NP  
; CURRENT APPLICATION NUMBER: US/10/651,237  
; CURRENT FILING DATE: 2003-08-27  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 1924  
; TYPE: DNA  
; ORGANISM: human  
US-10-651-237-4

Query Match 91.1% Score 16.4; DB 6; Length 1924;  
Best Local Similarity 94.4%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGGCCGCTGCTCTTG 18  
DB 1412 AGCAGGCCGCTGCTCTTG 1395

RESULT 2  
US-10-651-237-7/c  
; Sequence 7, Application US/10651237  
; GENERAL INFORMATION:  
; APPLICANT: Ortho-Clinical Diagnostics, Inc.  
; APPLICANT: Wang, Yixin  
; TITLE OF INVENTION: Colorectal Cancer Prognostics  
; FILE REFERENCE: ADS-5003 US NP  
; CURRENT APPLICATION NUMBER: US/10/651,237  
; CURRENT FILING DATE: 2003-08-27  
; NUMBER OF SEQ ID NOS: 94  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 1924  
; TYPE: DNA  
; ORGANISM: human  
US-10-651-237-7

Query Match 91.1% Score 16.4; DB 6; Length 1924;  
Best Local Similarity 94.4%; Pred. No. 65;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGCAGCCCGCTGTCCTTG 18  
|||  
Db 1412 AGCAGCCCGCTGTCCTTG 1395

RESULT 3  
US-60-495-114-38730/c

/ Sequence 38730, Application US/60495114  
/ GENERAL INFORMATION:  
/ APPLICANT: CARGILL, Michele  
/ TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
/ TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
/ FILE REFERENCE: CL001480  
/ CURRENT APPLICATION NUMBER: US/60/495,114  
/ CURRENT FILING DATE: 2003-08-15  
/ NUMBER OF SEQ ID NOS: 91238  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 38730  
/ LENGTH: 201  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-60-495-114-38730

Query Match 85.6%; Score 15.4; DB 7; Length 201;  
Best Local Similarity 94.1%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGCAGCCCGCTGTCCTT 17  
|||  
Db 92 AGCAGCCCGCTGTCCTT 76

RESULT 4  
US-60-495-114-89612/c

/ Sequence 89612, Application US/60495114  
/ GENERAL INFORMATION:  
/ APPLICANT: CARGILL, Michele  
/ TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
/ TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
/ FILE REFERENCE: CL001480  
/ CURRENT APPLICATION NUMBER: US/60/495,114  
/ CURRENT FILING DATE: 2003-08-15  
/ NUMBER OF SEQ ID NOS: 91238  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 89612  
/ LENGTH: 201  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-60-495-114-89612

Query Match 85.6%; Score 15.4; DB 7; Length 201;  
Best Local Similarity 94.1%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGCAGCCCGCTGTCCTT 17  
|||  
Db 64 AGCAGCCCGCTGTCCTT 48

RESULT 5  
US-60-495-114-89619/c

/ Sequence 89619, Application US/60495114  
/ GENERAL INFORMATION:  
/ APPLICANT: CARGILL, Michele  
/ TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
/ TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
/ FILE REFERENCE: CL001480  
/ CURRENT APPLICATION NUMBER: US/60/495,114  
/ CURRENT FILING DATE: 2003-08-15  
/ NUMBER OF SEQ ID NOS: 91238

/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 89619  
/ LENGTH: 201  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-60-495-114-89619

Query Match 85.6%; Score 15.4; DB 7; Length 201;  
Best Local Similarity 94.1%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGCAGCCCGCTGTCCTT 17  
|||  
Db 92 AGCAGCCCGCTGTCCTT 76

RESULT 6  
US-10-425-114A-26835

/ Sequence 26835, Application US/10425114A  
/ GENERAL INFORMATION:  
/ APPLICANT: Liu, Jindong  
/ APPLICANT: Zhou, Yihua  
/ APPLICANT: Kovalic, David K.  
/ APPLICANT: Screen, Steven E  
/ APPLICANT: Tabaska, Jack E  
/ APPLICANT: Cao, Yongwei  
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
/ FILE REFERENCE: 38-21(53313)B  
/ CURRENT APPLICATION NUMBER: US/10/425,114A  
/ CURRENT FILING DATE: 2003-04-28  
/ NUMBER OF SEQ ID NOS: 73128  
/ SEQ ID NO 26835  
/ LENGTH: 2441  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ OTHER INFORMATION: Clone ID: LIB4649-035-D9\_FLI  
US-10-425-114A-26835

Query Match 85.6%; Score 15.4; DB 6; Length 2441;  
Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGCAGCCCGCTGTCCTT 17  
|||  
Db 49 AGCAGCCCGCTGTCCTT 65

RESULT 7  
US-10-425-114A-26959

/ Sequence 26959, Application US/10425114A  
/ GENERAL INFORMATION:  
/ APPLICANT: Liu, Jindong  
/ APPLICANT: Zhou, Yihua  
/ APPLICANT: Kovalic, David K.  
/ APPLICANT: Screen, Steven E  
/ APPLICANT: Tabaska, Jack E  
/ APPLICANT: Cao, Yongwei  
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
/ FILE REFERENCE: 38-21(53313)B  
/ CURRENT APPLICATION NUMBER: US/10/425,114A  
/ CURRENT FILING DATE: 2003-04-28  
/ NUMBER OF SEQ ID NOS: 73128  
/ SEQ ID NO 26959  
/ LENGTH: 2463  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ OTHER INFORMATION: Clone ID: LIB4676-059-G1\_FLI  
US-10-425-114A-26959

Query Match 85.6%; Score 15.4; DB 6; Length 2463;  
 Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTT 17  
 |||||  
 DB 60 AGCAGCGCGTGTCTT 76

## RESULT 8

US-60-490-890-1861  
 ; Sequence 1861, Application US/60490890  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Li, Marcha  
 ; APPLICANT: Rupnow, Brent A.  
 ; APPLICANT: Webster, Kevin R.  
 ; APPLICANT: Jackson, Donald  
 ; APPLICANT: Wong, Tai W.  
 ; TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION  
 ; FILE REFERENCE: D0310 PSP  
 ; CURRENT APPLICATION NUMBER: US/60/490,890  
 ; CURRENT FILING DATE: 2003-07-29  
 ; NUMBER OF SEQ ID NOS: 2779  
 ; SOFTWARE: Patent version 3.2  
 ; SEQ ID NO 1861  
 ; LENGTH: 2470  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-60-490-890-1861

Query Match 85.6%; Score 15.4; DB 7; Length 2470;  
 Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTT 17  
 |||||  
 DB 78 AGCAGCGCGTGTCTT 94

## RESULT 9

US-10-425-114A-26285  
 ; Sequence 26285, Application US/10425114A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovacic, David K.  
 ; APPLICANT: Screen, Steven B.  
 ; APPLICANT: Tabaska, Jack E.  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53333)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114A  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 26285  
 ; LENGTH: 2487  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURES:  
 ; OTHER INFORMATION: Clone ID: LIB4119-174-G9\_FLI  
 US-10-425-114A-26285

Query Match 85.6%; Score 15.4; DB 6; Length 2487;  
 Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTT 17  
 |||||  
 DB 95 AGCAGCGCGTGTCTT 111

## RESULT 10

US-09-935-368A-2/c  
 ; Sequence 2, Application US/09935368A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cooper, Mark J.  
 ; TITLE OF INVENTION: Expression System for Production of  
 ; TITLE OF INVENTION: Therapeutic Proteins  
 ; FILE REFERENCE: 003659.00010  
 ; CURRENT APPLICATION NUMBER: US/09/935,368A  
 ; CURRENT FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 09/473,646  
 ; PRIOR FILING DATE: 1999-12-28  
 ; PRIOR APPLICATION NUMBER: PCT/US98/12777  
 ; PRIOR FILING DATE: 1998-06-19  
 ; PRIOR APPLICATION NUMBER: US 60/050,356  
 ; PRIOR FILING DATE: 1997-06-20  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2799  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-935-368A-2

Query Match 85.6%; Score 15.4; DB 5; Length 2799;  
 Best Local Similarity 94.1%; Pred. No. 1.9e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTT 17  
 |||||  
 DB 332 AGCAGCGCGTGTCTT 316

## RESULT 11

US-60-495-114-16387/c  
 ; Sequence 16387, Application US/60495114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele  
 ; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
 ; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; FILE REFERENCE: CLO01480  
 ; CURRENT APPLICATION NUMBER: US/60/495,114  
 ; CURRENT FILING DATE: 2003-08-15  
 ; NUMBER OF SEQ ID NOS: 91238  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16387  
 ; LENGTH: 26433  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURES:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(26433)  
 ; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-  
 US-60-495-114-16387

Query Match 85.6%; Score 15.4; DB 7; Length 26433;  
 Best Local Similarity 94.1%; Pred. No. 1.7e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTT 17  
 |||||  
 DB 1366 AGCAGCGCGTGTCTT 1350

## RESULT 12

US-60-495-114-16957/c  
 ; Sequence 16957, Application US/60495114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele  
 ; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES  
 ; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; FILE REFERENCE: CLO01480

```
/ CURRENT APPLICATION NUMBER: US/60/495,114
/ CURRENT FILING DATE: 2003-08-15
/ NUMBER OF SEQ ID NOS: 91238
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 16957
/ LENGTH: 45664
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(45664)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-495-114-16957
```

```
Query Match      85.6%; Score 15.4; DB 7; Length 45664;
Best Local Similarity 94.1%; Pred. No. 1.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCAGGCCGCTGTCCTT 17
Db      37645 AGCAGGCCGCTGTCCTT 37629
```

```
RESULT 13
US-60-495-114-16382/C
/ Sequence 16382, Application US/60495114
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele
/ TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
/ TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
/ FILE REFERENCE: CL001480
/ CURRENT APPLICATION NUMBER: US/60/495,114
/ CURRENT FILING DATE: 2003-08-15
/ NUMBER OF SEQ ID NOS: 91238
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 16382
/ LENGTH: 384485
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(384485)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-495-114-16382
```

```
Query Match      85.6%; Score 15.4; DB 7; Length 384485;
Best Local Similarity 94.1%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCAGGCCGCTGTCCTT 17
Db      376152 AACAGGCCGCTGTCCTT 376136
```

```
RESULT 14
US-10-425-114A-27906
/ Sequence 27906, Application US/10425114A
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovacic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(5313)B
/ CURRENT APPLICATION NUMBER: US/10/425,114A
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 27906
/ LENGTH: 1529
```

```
/ TYPE: DNA
/ ORGANISM: Zea mays
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB4742-003-G8_FLI
US-10-425-114A-27906
```

```
Query Match      83.3%; Score 15; DB 6; Length 1529;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 GCAGGCCGCTGTCCTT 16
Db      127 GCAGGCCGCTGTCCTT 141
```

```
RESULT 15
US-60-495-114-15241/C
/ Sequence 15241, Application US/60495114
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele
/ TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
/ TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
/ FILE REFERENCE: CL001480
/ CURRENT APPLICATION NUMBER: US/60/495,114
/ CURRENT FILING DATE: 2003-08-15
/ NUMBER OF SEQ ID NOS: 91238
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 15241
/ LENGTH: 201
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-60-495-114-15241
```

```
Query Match      82.2%; Score 14.8; DB 7; Length 201;
Best Local Similarity 88.9%; Pred. No. 4.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 AGCAGGCCGCTGTCCTTG 18
Db      65 AGCAGGCCGCTGTCCTTG 48
```

```
Search completed: September 16, 2003, 22:47:53
Job time : 30.3366 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:28:15 ; Search time 2175.74 Seconds

(without alignments)  
457.275 Million cell updates/sec

Title: US-09-594-065-3

Perfect score: 30  
1 cccctgctagctgctagcactctctgcagcg 30

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 33363688 seqs, 1658189874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_NA\_Main:\*

```

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US092A_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US092B_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US093A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US098D_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*

```

```

44: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US103A_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US103B_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US104A_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US104B_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US6047_COMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	25	US-09-594-065-3
2	23.8	79.3	2142	25	US-09-584-852-6658
3	20.4	68.0	714	51	US-10-369-453-41285
4	20.4	68.0	714	91	US-60-360-039-41285

5 20.4 68.0 2731748 49 US-10-297-465A-1 Sequence 1, Appl1  
6 20.2 67.7 372 18 US-09-270-849B-77960 Sequence 77960, A  
7 19.4 64.7 376 18 US-09-270-849B-51162 Sequence 51162, A  
8 19.4 64.7 483 51 US-10-366-683-6704 Sequence 6704, Ap  
9 19.4 64.7 483 52 US-10-366-683-6704 Sequence 6704, Ap  
10 19.4 64.7 490 18 US-09-270-849B-169678 Sequence 169678, A  
11 19.4 64.7 510 51 US-10-366-683-6599 Sequence 6599, Ap  
12 19.4 64.7 510 52 US-10-419-128-6599 Sequence 6599, Ap  
13 19.4 64.7 1302 49 US-10-282-122A-23788 Sequence 23788, A  
14 19.4 64.7 1323 47 US-10-156-761-4750 Sequence 4750, Ap  
15 19.4 64.7 1368 21 US-09-489-039A-6204 Sequence 6204, Ap  
16 19.4 64.7 1368 53 US-10-446-203-6204 Sequence 6204, Ap  
17 19.4 64.7 2022 49 US-10-263-929-79 Sequence 79, Appl1  
18 19.4 64.7 2079 46 US-10-144-771-8606 Sequence 8606, Ap  
19 19.4 64.7 2079 91 US-09-360-207-8606 Sequence 8606, Ap  
20 19.4 64.7 3084 51 US-10-366-683-6539 Sequence 6539, Ap  
21 19.4 64.7 3084 52 US-10-419-128-6539 Sequence 6539, Ap  
22 19.4 64.7 3273 51 US-10-366-683-6578 Sequence 6578, Ap  
23 19.4 64.7 3273 52 US-10-419-128-6578 Sequence 6578, Ap  
24 19.4 64.7 110079 1 PCT-US02-19457-96 Sequence 96, Appl1  
25 19.4 64.7 110079 47 US-10-175-523-96 Sequence 96, Appl1  
26 19.4 64.7 9025608 47 US-10-156-761-1 Sequence 1, Appl1  
27 19.2 64.0 106174 79 US-09-248-505-419 Sequence 419, Appl1  
28 19.2 63.3 482 18 US-09-270-849B-65208 Sequence 65208, A  
29 19.2 63.3 487 22 US-09-521-640-19436 Sequence 19436, A  
30 19.2 63.3 528 18 US-09-270-849B-43567 Sequence 43567, A  
31 19.2 63.3 1485 53 US-10-437-963-102206 Sequence 102206, A  
32 19.2 63.3 1485 53 US-10-438-246-11944 Sequence 11944, A  
33 19.2 63.3 1485 53 US-10-438-246-21582 Sequence 21582, A  
34 18.8 62.7 444 26 US-09-606-977-56997 Sequence 56997, A  
35 18.8 62.7 444 68 US-09-141-233-56997 Sequence 56997, A  
36 18.8 62.7 576 32 US-09-738-626-2147 Sequence 2147, A  
37 18.8 62.7 578 72 US-09-182-316-4088 Sequence 4088, Ap  
38 18.8 62.7 726 26 US-09-605-703B-1951 Sequence 1951, Ap  
39 18.8 62.7 1368 16 US-09-191-989-3 Sequence 3, Appl1  
40 18.8 62.7 1593 47 US-10-156-761-5071 Sequence 5071, Ap  
41 18.8 62.7 2936 16 US-09-191-989-1 Sequence 1, Appl1  
42 18.8 62.7 4454 16 US-09-191-989-2 Sequence 2, Appl1  
43 18.8 62.7 99916 34 US-09-816-095-3 Sequence 3, Appl1  
44 18.8 62.7 99916 53 US-10-634-905-3 Sequence 3, Appl1  
45 18.8 62.7 113462 22 US-09-528-237A-1860 Sequence 1860, Ap

## ALIGNMENTS

RESULT 1  
US-09-594-065-3  
; Sequence 3, Application US/09594065  
; GENERAL INFORMATION:  
; APPLICANT: WalkerPeach, Cindy  
; APPLICANT: DuBois, Dwight  
; TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an  
; FILE REFERENCE: 25436/1280  
; CURRENT APPLICATION NUMBER: US/09/594,065  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: IAC specific molecular beacon  
US-09-594-065-3

Query Match 100.0%; Score 30; DB 25; Length 30;  
Best Local Similarity 100.0%; Pred. No. 0.011; Mismatches 0; Indels 0; Gaps 0;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 CCTGCGTAGTGTACGACCTCTCGAGGG 30  
|||||

Db 1 CCTGCGTAGTGTACGACCTCTCGAGGG 30

RESULT 2  
US-09-584-852-6658  
; Sequence 6658, Application US/09584852  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; APPLICANT: Williamson, Mark  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES  
; FILE REFERENCE: 1600.1118-001  
; CURRENT APPLICATION NUMBER: US/09/584,852  
; CURRENT FILING DATE: 2000-05-24  
; PRIOR APPLICATION NUMBER: 60/135,618  
; PRIOR FILING DATE: 1999-05-24  
; PRIOR APPLICATION NUMBER: 60/135,711  
; PRIOR FILING DATE: 1999-05-24  
; PRIOR APPLICATION NUMBER: 60/135,627  
; PRIOR FILING DATE: 1999-05-24  
; NUMBER OF SEQ ID NOS: 8040  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6658  
; LENGTH: 2142  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-584-852-6658

Query Match 79.3%; Score 23.8; DB 25; Length 2142;  
Best Local Similarity 92.6%; Pred. No. 8.9;  
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 2 CCTGCGTAGTGTACGACCTCTCGAG 28  
446 CCTGCGTAGTGTACGACCTCTCGAG 472

RESULT 3  
US-10-369-493-41285/C  
; Sequence 41285, Application US/10369493  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 41285  
; LENGTH: 714  
; TYPE: DNA  
; ORGANISM: Xylella fastidiosa  
US-10-369-493-41285

Query Match 68.0%; Score 20.4; DB 51; Length 714;  
Best Local Similarity 80.0%; Pred. No. 2.5e+02;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 1 CCTGCGTAGTGTACGACCTCTCGAGGG 30  
403 CCTGCGTAGTGTACGACCTCTCGAGGG 374

RESULT 4  
US-60-360-039-41285/C  
; Sequence 41285, Application US/60360039  
; GENERAL INFORMATION:



RESULT 9  
US-10-419-128-6704/c  
Sequence 6704, Application US/10419128  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/10/419,128  
CURRENT FILING DATE: 2003-04-21  
PRIOR APPLICATION NUMBER: US/09/252,991  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6704  
LENGTH: 483  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-419-128-6704

Query Match 64.7%; Score 19.4; DB 52; Length 483;  
Best Local Similarity 79.3%; Pred. No. 6.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCTGCGTAGTGTGACGACCTCTGTCAGG 29  
DB 35 CCCGGCGAGTGGGCGAGCGCTGTCAGG 7

RESULT 10  
US-09-270-849B-169678/c  
Sequence 169678, Application US/09270849B  
GENERAL INFORMATION:  
APPLICANT: Swimmer et al.  
TITLE OF INVENTION: Insect genome survey devices  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/09/270,849B  
CURRENT FILING DATE: 1999-03-17  
NUMBER OF SEQ ID NOS: 195450  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 169678  
LENGTH: 490  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-270-849B-169678

Query Match 64.7%; Score 19.4; DB 18; Length 490;  
Best Local Similarity 79.3%; Pred. No. 6.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CCTGCGTAGTGTGACGACCTCTGTCAGG 30  
DB 244 CCTGCGTAGTGTGACGACCTCTGTCAGG 216

RESULT 11  
US-10-366-683-6599  
Sequence 6599, Application US/10366683  
GENERAL INFORMATION:  
APPLICANT: Rubenfield, Marc J.  
APPLICANT: Nolling, Jock  
APPLICANT: Deloughery, Craig  
APPLICANT: Bush, David  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: PATH03-04  
CURRENT APPLICATION NUMBER: US/10/366,683

CURRENT FILING DATE: 2003-02-13  
PRIOR APPLICATION NUMBER: 09/252,991  
PRIOR FILING DATE: 1999-02-18  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6599  
LENGTH: 510  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-366-683-6599

Query Match 64.7%; Score 19.4; DB 51; Length 510;  
Best Local Similarity 79.3%; Pred. No. 6.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCTGCGTAGTGTGACGACCTCTGTCAGG 29  
DB 347 CCCGGCGAGTGGGCGAGCGCTGTCAGG 375

RESULT 12  
US-10-419-128-6599  
Sequence 6599, Application US/10419128  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/10/419,128  
CURRENT FILING DATE: 2003-04-21  
PRIOR APPLICATION NUMBER: US/09/252,991  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 6599  
LENGTH: 510  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-419-128-6599

Query Match 64.7%; Score 19.4; DB 52; Length 510;  
Best Local Similarity 79.3%; Pred. No. 6.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCTGCGTAGTGTGACGACCTCTGTCAGG 29  
DB 347 CCCGGCGAGTGGGCGAGCGCTGTCAGG 375

RESULT 13  
US-10-282-122A-23788  
Sequence 23788, Application US/10282122A  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Habelbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA 034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 23788  
LENGTH: 1302  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-10-282-122A-23788

Query Match 64.7% Score 19.4; DB 49; Length 1302;  
Best Local Similarity 79.3%; Pred. No. 7.3e+02;

Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTGACGACCTCTGCAGG 29  
Db 708 CCCGCGTGTGTGACGAGGTTCTCTGCAGG 736

RESULT 14  
US-10-156-761-4750  
Sequence 4750, Application US/10156761  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 4750  
LENGTH: 1323  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1323)  
US-10-156-761-4750

Query Match 64.7% Score 19.4; DB 47; Length 1323;  
Best Local Similarity 79.3%; Pred. No. 7.3e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTGACGACCTCTGCAGG 29  
Db 669 CCCGCGAAGCGGTACGAGCTCTCTGCAGG 697

RESULT 15  
US-09-489-039A-6204  
Sequence 6204, Application US/09489039A  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 6204  
LENGTH: 1368  
TYPE: DNA  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-6204

Query Match 64.7% Score 19.4; DB 21; Length 1368;  
Best Local Similarity 79.3%; Pred. No. 7.4e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CCTGCGTAGTGTGACGACCTCTGCAGG 29  
Db 771 CCCGCGTGTGTGACGAGGTTCTCTGCAGG 799

Search completed: September 16, 2003, 22:44:55  
Job time : 2185.74 secs

**BEST AVAILABLE COPY**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:37:15 ; Search time 47.2277 Seconds  
(without alignments)  
535.803 Million cell updates/sec

Title: US-09-594-065-3

Perfect score: 30

Sequence: 1 cccgcgcagtgatgcacccctcgcaggg 30

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 606776 seqs, 421745864 residues

Total number of hits satisfying chosen parameters: 1213552

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents\_MA\_New.\*  
1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB\_seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB\_seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB\_seq.\*  
4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB\_seq.\*  
5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB\_seq.\*  
6: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB\_seq.\*  
7: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB\_seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.4	68.0	2731748	6 US-10-297-465B-1	Sequence 1, Appl
2	18.8	62.7	4813087	5 US-09-947-914-75	Sequence 75, Appl
3	18	60.0	1627	6 US-10-425-114A-33391	Sequence 33391, A
4	17.8	59.3	1189	6 US-10-425-114A-3571	Sequence 3571, Ap
5	17.8	59.3	1268	6 US-10-425-114A-19153	Sequence 19153, A
6	17.8	59.3	1327	6 US-10-425-114A-5952	Sequence 5952, Ap
7	17.8	59.3	1746	6 US-10-425-114A-28420	Sequence 28420, A
8	17.6	58.7	1111	6 US-10-425-114A-1209	Sequence 1209, Ap
9	17.6	58.7	4829	6 US-10-428-487-41	Sequence 41, Appl
10	17.4	58.0	263	1 PCT-US03-18714-21709	Sequence 21709, A
11	17.4	58.0	734	7 US-60-480-590-5114	Sequence 5114, Ap
12	17.4	58.0	1445	6 US-10-350-923B-49	Sequence 49, Appl
13	17.4	58.0	1647	6 US-10-425-114A-1130	Sequence 1130, Ap
14	17.4	58.0	100537	7 US-60-485-450-11874	Sequence 11874, A
15	17.4	58.0	4813087	5 US-09-947-914-75	Sequence 75, Appl
16	17.2	57.3	201	7 US-60-487-610-31672	Sequence 31672, A
17	17.2	57.3	201	7 US-60-487-610-31706	Sequence 31706, A
18	17.2	57.3	201	7 US-60-487-610-35125	Sequence 35125, A
19	17.2	57.3	201	7 US-60-485-450-5775	Sequence 5775, Ap
20	17.2	57.3	201	7 US-60-485-450-26392	Sequence 26392, A
21	17.2	57.3	905	6 US-10-425-114A-13077	Sequence 13077, A
22	17.2	57.3	974	6 US-10-425-114A-14815	Sequence 14815, A
23	17.2	57.3	1121	1 PCT-US03-20480-60	Sequence 60, Appl
24	17.2	57.3	4800	7 US-60-485-450-375	Sequence 375, Appl
25	17.2	57.3	76269	7 US-60-485-450-12031	Sequence 12031, A
26	17.2	57.3	127226	7 US-60-487-610-19335	Sequence 19335, A

C	27	17.2	57.3	127313	7	US-60-487-610-19398	Sequence 19398, A
C	28	17	56.7	1224	6	US-10-296-115-154	Sequence 154, Appl
C	29	17	56.7	8059021	5	US-09-947-914-53	Sequence 53, Appl
C	30	16.8	56.0	201	7	US-60-487-610-57668	Sequence 57668, A
C	31	16.8	56.0	1981	7	US-60-487-610-793	Sequence 793, Appl
C	32	16.8	56.0	2017	6	US-10-296-115-464	Sequence 464, Appl
C	33	16.8	56.0	2052	7	US-60-493-007-2834	Sequence 2834, Ap
C	34	16.8	56.0	2168	5	US-09-908-576-331	Sequence 331, Appl
C	35	16.8	56.0	2189	7	US-60-487-610-789	Sequence 789, Appl
C	36	16.8	56.0	2196	7	US-60-487-610-790	Sequence 790, Appl
C	37	16.8	56.0	2320	7	US-60-487-610-788	Sequence 788, Appl
C	38	16.8	56.0	2366	7	US-60-487-610-794	Sequence 794, Appl
C	39	16.8	56.0	2473	7	US-60-487-610-795	Sequence 795, Appl
C	40	16.8	56.0	2657	7	US-60-487-610-791	Sequence 791, Appl
C	41	16.8	56.0	2663	7	US-60-487-610-792	Sequence 792, Appl
C	42	16.8	56.0	2715	7	US-60-487-610-796	Sequence 796, Appl
C	43	16.8	56.0	9041	7	US-60-478-196-332	Sequence 332, Appl
C	44	16.8	56.0	26562	7	US-60-487-610-20041	Sequence 20041, A
C	45	16.8	56.0	45317	7	US-60-487-610-19569	Sequence 19569, A

## ALIGNMENTS

```
RESULT 1
US-10-297-465B-1
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Setubal, Joao
; APPLICANT: Medeiros, Joao
; APPLICANT: Arruda, Paulo
; TITLE OF INVENTION: Isolated Genome of Xylella fastidiosa and Uses Thereof
; FILE REFERENCE: PAPER 202 US (10213376)
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: US/10/297,465B
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1
Query Match      68.0%; Score 20.4; DB 6; Length 2731748;
Beat Local Similarity 80.0%; Pred. No. 8.5;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY      1 cccgcgcagtgatgcacccctcgcaggg 30
        |||||  |||  |||||  |||||
DB      1446654 cccgcgcagtgatgcacccctcgcaggg 1446683
RESULT 2
US-09-947-914-75
; Sequence 75, Application US/09947914
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH
; FILE REFERENCE: CLO01298
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 75
; SEQ ID NO 75
; LENGTH: 4813087
; TYPE: DNA
; ORGANISM: HUMAN
```

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)... (4813087)  
OTHER INFORMATION: n = A,T,C or G  
US-09-947-914-75

Query Match  
Best Local Similarity 62.7%; Score 18.8; DB 5; Length 4813087;  
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CCTCGCTAGTGTAGTACGACTCTCTGCAGG 30  
DB 2035933 CCTGAGCATGTAGTACGCGCTCTGCAGG 2035962

RESULT 3  
US-10-425-114A-33391  
Sequence 33391, Application US/10425114A  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 33391  
LENGTH: 1627  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-2MFLMO17105A02\_FLI  
US-10-425-114A-33391

Query Match  
Best Local Similarity 60.0%; Score 18; DB 6; Length 1627;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CTGCGTAGTGTAGTACGACTCTCTGCAG 28  
DB 130 CGCGGAGGGGTTCCGACATCTCTGCAG 155

RESULT 4  
US-10-425-114A-3571/C  
Sequence 3571, Application US/10425114A  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 3571  
LENGTH: 1189  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700331949\_FLI  
US-10-425-114A-3571

Query Match  
Best Local Similarity 59.3%; Score 17.8; DB 6; Length 1189;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Best Local Similarity 75.9%; Pred. No. 87;  
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CCTGCGTAGTGTAGTACGACTCTCTGCAGG 30  
DB 1035 CCTATATAGTAGTATGACCACTCTGCAGG 1007

RESULT 5  
US-10-425-114A-19153  
Sequence 19153, Application US/10425114A  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 19153  
LENGTH: 1268  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3137-002-D11\_FLI  
US-10-425-114A-19153

Query Match  
Best Local Similarity 59.3%; Score 17.8; DB 6; Length 1268;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GTGTAGCAGACTCTCTGCAGG 30  
DB 207 GTGTAGCAGACTCTCTGCAGG 227

RESULT 6  
US-10-425-114A-5952  
Sequence 5952, Application US/10425114A  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114A  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 5952  
LENGTH: 1327  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700549530\_FLI  
US-10-425-114A-5952

Query Match  
Best Local Similarity 59.3%; Score 17.8; DB 6; Length 1327;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GTGTAGCAGACTCTCTGCAGG 30  
DB 209 GTGTAGCAGACTCTCTGCAGG 229



```
RESULT 7
US-10-425-114A-28420/C
; Sequence 28420, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 28420
; LENGTH: 1746
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4757-014-A3_FLI
US-10-425-114A-28420

Query Match          59.3%; Score 17.8; DB 6; Length 1746;
Best Local Similarity 75.9%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTTGGCTAGTGTAGACCTCTGCGAGG 30
Db 1640 CCTATATAGTAGTACGACCACTGCGAGG 1612

RESULT 8
US-10-425-114A-1209
; Sequence 1209, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1209
; LENGTH: 1111
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700104758_FLI
US-10-425-114A-1209

Query Match          58.7%; Score 17.6; DB 6; Length 1111;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 CGTAGGTAGACGACCTCTGCGAGG 29
Db 846 CCGCGTAGACCACTCTGCGAGG 869

RESULT 9
US-10-428-487-41
; Sequence 41, Application US/10428487
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA K.
; APPLICANT: GERBER, HANS-PETER
```

```
; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
; FILE REFERENCE: 09800080-0103
; CURRENT APPLICATION NUMBER: US/10/428,487
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: 09/815,153
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,201
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 4829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-428-487-41

Query Match          58.7%; Score 17.6; DB 6; Length 4829;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 CGTAGGTAGACGACCTCTGCGAGG 29
Db 125 CTTGGTAGACCACTCTGCGAGG 148

RESULT 10
PCT-US03-18714-21709/C
; Sequence 21709, Application PC/TUS0318714
; GENERAL INFORMATION:
; APPLICANT: Reguime Corporation
; TITLE OF INVENTION: Functional Sites
; FILE REFERENCE: 11207-029-228
; CURRENT APPLICATION NUMBER: PCT/US03/18714
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/387,887
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/387,910
; PRIOR FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 51999
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21709
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Homo sapiens
PCT-US03-18714-21709

Query Match          58.0%; Score 17.4; DB 1; Length 263;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 CTGCGTAGGTAGACGACCTCTGCGAGG 29
Db 151 CTGCGACTCGGAGGAACTCTGCGAGG 125

RESULT 11
US-60-480-590-5114/C
; Sequence 5114, Application US/60480590
; GENERAL INFORMATION:
; APPLICANT: Monsanto
; APPLICANT: Monsanto
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Soybean Polymorphisms and Methods of Genotyping
; FILE REFERENCE: 38-15 (53382)
; CURRENT APPLICATION NUMBER: US/60/480,590
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 6578
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5114
; LENGTH: 734
; TYPE: DNA
; ORGANISM: Glycine max
US-60-480-590-5114
```

Query Match 58.0%; Score 17.4; DB 7; Length 724;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CCTGGCTAGTGTACGACCTCTCGAG 28  
DB 43 CCTGGCTAGTGTGTGATCACCCTGAG 17

## RESULT 12

US-10-350-923B-49  
; Sequence 49, Application US/10350923B  
; GENERAL INFORMATION:  
; APPLICANT: Dotson, Stanton B.  
; APPLICANT: Ma, Xiao Jun  
; TITLE OF INVENTION: Nucleic acid, polypeptides, vectors, and cells derived from active  
; FILE REFERENCE: 503157-01  
; CURRENT APPLICATION NUMBER: US/10/350,923B  
; CURRENT FILING DATE: 2003-01-24  
; PRIOR APPLICATION NUMBER: US/ 09/454,280  
; PRIOR FILING DATE: 1999-12-03  
; PRIOR APPLICATION NUMBER: PCT/US99/28773  
; PRIOR FILING DATE: 1999-06-12  
; PRIOR APPLICATION NUMBER: US 60/111,006  
; PRIOR FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 49  
; LENGTH: 1445  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-350-923B-49

Query Match 58.0%; Score 17.4; DB 6; Length 1445;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CCTGGCTAGTGTACGACCTCTCGAG 28  
DB 226 CCGAGTCTCTGTACGACCTCTCGAG 252

## RESULT 13

US-10-425-114A-3130/C  
; Sequence 3130, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jindong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 3130  
; LENGTH: 1647  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700241792\_FLI  
US-10-425-114A-3130

Query Match 58.0%; Score 17.4; DB 6; Length 1647;  
Best Local Similarity 77.8%; Pred. No. 1.3e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 CTGCTAGTGTGTACGACCTCTCGAG 29

DB 867 CTGCGAAGGCTTCCAACTCTCGAGG 841

## RESULT 14

US-60-485-450-11874  
; Sequence 11874, Application US/60485450  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: CHANG, Sheng-Yung  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
; FILE REFERENCE: CLO01470  
; CURRENT APPLICATION NUMBER: US/60/485,450  
; CURRENT FILING DATE: 2003-07-09  
; NUMBER OF SEQ ID NOS: 47859  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11874  
; LENGTH: 100537  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-60-485-450-11874

Query Match 58.0%; Score 17.4; DB 7; Length 100537;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TGGCTAGTGTACGACCTCTCGAGG 30  
DB 6578 TGTGTAAGGTACTCTCTCTCGAG 6604

## RESULT 15

US-09-947-914-75/C  
; Sequence 75, Application US/09947914  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig  
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH  
; FILE REFERENCE: CLO01298  
; CURRENT APPLICATION NUMBER: US/09/947,914  
; CURRENT FILING DATE: 2001-09-07  
; NUMBER OF SEQ ID NOS: 75  
; SEQ ID NO 75  
; LENGTH: 4813087  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(4813087)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-947-914-75

Query Match 58.0%; Score 17.4; DB 5; Length 4813087;  
Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 CCTGGCTAGTGTACGACCTCTCGA 27  
DB 1761296 CCTGGCTGGGGAATATCTCTCGA 1761270

Search completed: September 16, 2003, 22:48:13  
Job time : 67.2277 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:28:15 ; Search time 2393.32 Seconds  
(without alignments)  
457.275 Million cell updates/sec

Title: US-09-594-065-4

Perfect score: 33  
Sequence: 1 cccctgaactcgtgtccctccagcagcaggg 33

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3363688 seqs, 1658189874 residues

Total number of hits satisfying chosen parameters: 66727376

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_MA Main:\*

```

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*old:*
3: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*

```

```

44: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32.6	98.8	33	US-09-594-065-4	Sequence 4, Appli
2	32.6	98.8	33	US-09-594-065-10	Sequence 10, Appli
3	31.4	95.2	33	US-09-594-065-11	Sequence 11, Appli
4	22.6	68.5	39	US-09-594-065-7	Sequence 7, Appli

5 22.6 68.5 109 25 US-09-594-065-13 Sequence 13, Appl  
6 22.6 68.5 2645 6 US-08-123-456-141 Sequence 141, App  
7 22.6 68.5 2645 18 US-09-129-477A-141 Sequence 141, App  
8 22.6 68.5 2645 43 US-09-994-404-141 Sequence 141, App  
9 22.6 68.5 2641 6 US-08-123-456-78 Sequence 78, Appl  
10 22.6 68.5 2641 18 US-09-129-477A-78 Sequence 78, Appl  
11 22.6 68.5 2641 43 US-09-994-404-78 Sequence 78, Appl  
12 22.6 68.5 117213 6 US-08-123-456-217 Sequence 217, Appl  
13 22.6 68.5 117213 18 US-09-129-477A-217 Sequence 217, Appl  
14 22.6 68.5 117213 43 US-09-994-404-217 Sequence 217, Appl  
15 22.6 68.5 154746 1 PCR-US01-11372-8 Sequence 8, Appl  
16 22.6 68.5 154746 2 PCR-US01-11372-8 Sequence 8, Appl  
17 22.6 68.5 154746 3 PCR-US01-11372-8 Sequence 8, Appl  
18 22.6 68.5 154746 15 US-09-065-511-3362 Sequence 3362, Ap  
19 22.6 67.9 188 23 US-09-539-331D-25769 Sequence 25769, A  
20 22.4 67.9 201 100 US-09-452-680-72811 Sequence 72811, A  
21 22.4 67.9 247 14 US-08-937-141-1744 Sequence 1744, Ap  
22 22.4 67.9 247 24 US-09-540-229-75778 Sequence 75778, A  
23 22.4 67.9 247 56 US-09-027-778-1744 Sequence 1744, Ap  
24 22.4 67.9 257 15 US-09-076-667-1489 Sequence 1489, Ap  
25 22.4 67.9 257 24 US-09-540-229-89186 Sequence 89186, A  
26 22.4 67.9 257 58 US-09-048-002-1489 Sequence 1489, Ap  
27 22.4 67.9 259 24 US-09-540-229-100072 Sequence 100072, A  
28 22.4 67.9 271 16 US-09-105-427-127 Sequence 127, App  
29 22.4 67.9 271 24 US-09-540-208-1734 Sequence 1734, Ap  
30 22.4 67.9 294 24 US-09-540-764-49141 Sequence 49141, A  
31 22.4 67.9 294 50 US-10-349-781-49141 Sequence 49141, A  
32 22.4 67.9 319 15 US-09-079-506-42 Sequence 42, Appl  
33 22.4 67.9 319 24 US-09-540-208-5526 Sequence 5526, Ap  
34 22.4 67.9 478 22 US-09-528-409-21764 Sequence 21764, A  
35 22.4 67.9 478 39 US-09-933-524-21764 Sequence 21764, A  
36 22.4 67.9 478 39 US-09-933-524-21764 Sequence 21764, A  
37 22.4 67.9 487 18 US-09-277-227-18868 Sequence 18868, A  
38 22.4 67.9 487 19 US-09-346-956-16027 Sequence 16027, A  
39 22.4 67.9 487 38 US-09-904-703-16027 Sequence 16027, A  
40 22.4 67.9 487 38 US-09-903-627-18868 Sequence 18868, A  
41 22.4 67.9 502 44 US-10-023-386-11498 Sequence 11498, A  
42 22.4 67.9 551 44 US-10-023-386-25205 Sequence 25205, A  
43 22.4 67.9 573 24 US-09-540-212A-11467 Sequence 61467, A  
44 22.4 67.9 601 40 US-09-947-907-2022 Sequence 2022, Ap  
45 22.4 67.9 601 40 US-09-949-016-37689 Sequence 37689, A

## ALIGNMENTS

RESULT 1  
US-09-594-065-4 Application US/09594065  
GENERAL INFORMATION:  
APPLICANT: Walkerpach, Cindy  
APPLICANT: Dubois, Dwight  
TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an  
FILE REFERENCE: 25436/1280  
CURRENT APPLICATION NUMBER: US/09/594,065  
CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 4  
LENGTH: 33  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: HSV-type common (TC) molecular beacon  
US-09-594-065-4

Query Match 98.8%; Score 32.6; DB 25; Length 33;  
Best Local Similarity 100.0%; Pred. No. 0.0083;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CCTGCAAACTGCTGCTCTCCAGCATGCAGG 33  
RESULT 2  
US-09-594-065-10  
GENERAL INFORMATION:  
APPLICANT: Walkerpach, Cindy  
APPLICANT: Dubois, Dwight  
TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an  
FILE REFERENCE: 25436/1280  
CURRENT APPLICATION NUMBER: US/09/594,065  
CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 10  
LENGTH: 33  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: HSV type-1 specific molecular beacon  
US-09-594-065-10

Query Match 98.8%; Score 32.6; DB 25; Length 33;  
Best Local Similarity 97.0%; Pred. No. 0.0083;  
Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 CCTGCAAACTGCTGCTCTCCAGCATGCAGG 33  
RESULT 3  
US-09-594-065-11  
GENERAL INFORMATION:  
APPLICANT: Walkerpach, Cindy  
APPLICANT: Dubois, Dwight  
TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an  
FILE REFERENCE: 25436/1280  
CURRENT APPLICATION NUMBER: US/09/594,065  
CURRENT FILING DATE: 2000-06-14  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 11  
LENGTH: 33  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: HSV type-2 specific molecular beacon  
US-09-594-065-11

Query Match 95.2%; Score 31.4; DB 25; Length 33;  
Best Local Similarity 97.0%; Pred. No. 0.026;  
Matches 32; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CCTGCAAACTGCTGCTCTCCAGCATGCAGG 33  
RESULT 4  
US-09-594-065-7/c  
GENERAL INFORMATION:  
APPLICANT: Walkerpach, Cindy  
APPLICANT: Dubois, Dwight  
TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an  
FILE REFERENCE: 25436/1280  
CURRENT APPLICATION NUMBER: US/09/594,065

```
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 39
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Single stranded oligonucleotide molecular beacon target for HSV 1
US-09-594-065-7
```

```
Query Match      68.5%; Score 22.6; DB 25; Length 39;
Best Local Similarity 95.7%; Pred. No. 1.1e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      6 CAAACTCGTGKCTCCGACGATG 28
Db      31 CAAACTCGTGKCTCCGACGATG 9
```

```
RESULT 5
US-09-594-065-13/c
; Sequence 13, Application US/09594065
; GENERAL INFORMATION:
; APPLICANT: WalkerPeach, Cindy
; APPLICANT: Dubois, Dwight
; TITLE OF INVENTION: Compositions, methods and kits for Herpes Simplex Virus Type 1 an
; FILE REFERENCE: 25436/1280
; CURRENT APPLICATION NUMBER: US/09/594,065
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 109
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Sequence amplified by PCR for HSV type-2
US-09-594-065-13
```

```
Query Match      68.5%; Score 22.6; DB 25; Length 109;
Best Local Similarity 95.7%; Pred. No. 1.2e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      6 CAAACTCGTGKCTCCGACGATG 28
Db      60 CAAACTCGTGKCTCCGACGATG 38
```

```
RESULT 6
US-08-123-456-141
; Sequence 141, Application US/08123456
; GENERAL INFORMATION:
; APPLICANT: ESSER, KLAUS M.
; APPLICANT: CHAN, JOHN Y.
; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
; APPLICANT: DELVECCHIO, ALFRED MICHAEL
; APPLICANT: DILLON, SUSAN B.
; APPLICANT: LEARY, JEFFREY JOSEPH
; APPLICANT: SUTTON, DAVID
; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
; FILE REFERENCE: P50583
; CURRENT APPLICATION NUMBER: US/08/123,456
; CURRENT FILING DATE: 1997-03-03
; PRIOR APPLICATION NUMBER: US 60/030,279
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/049,018
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 303
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 141
```

```
; LENGTH: 2645
; TYPE: DNA
; ORGANISM: Herpes simplex
US-08-123-456-141
```

```
Query Match      68.5%; Score 22.6; DB 6; Length 2645;
Best Local Similarity 95.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      6 CAAACTCGTGKCTCCGACGATG 28
Db      755 CAAACTCGTGKCTCCGACGATG 777
```

```
RESULT 7
US-09-297-477A-141
; Sequence 141, Application US/09297477A
; GENERAL INFORMATION:
; APPLICANT: ESSER, KLAUS M.
; APPLICANT: CHAN, JOHN Y.
; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
; APPLICANT: DELVECCHIO, ALFRED MICHAEL
; APPLICANT: DILLON, SUSAN B.
; APPLICANT: LEARY, JEFFREY JOSEPH
; APPLICANT: SUTTON, DAVID
; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
; FILE REFERENCE: P50583
; CURRENT APPLICATION NUMBER: US/09/297,477A
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/030,279
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/049,018
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 303
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 141
; LENGTH: 2645
; TYPE: DNA
; ORGANISM: Herpes simplex
US-09-297-477A-141
```

```
Query Match      68.5%; Score 22.6; DB 18; Length 2645;
Best Local Similarity 95.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      6 CAAACTCGTGKCTCCGACGATG 28
Db      755 CAAACTCGTGKCTCCGACGATG 777
```

```
RESULT 8
US-09-994-404-141
; Sequence 141, Application US/09994404
; GENERAL INFORMATION:
; APPLICANT: ESSER, KLAUS M.
; APPLICANT: CHAN, JOHN Y.
; APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
; APPLICANT: DELVECCHIO, ALFRED MICHAEL
; APPLICANT: DILLON, SUSAN B.
; APPLICANT: LEARY, JEFFREY JOSEPH
; APPLICANT: SUTTON, DAVID
; TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
; FILE REFERENCE: P50583
; CURRENT APPLICATION NUMBER: US/09/994,404
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 09/297,477
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: US 60/030,279
; PRIOR FILING DATE: 1999-11-04
; PRIOR APPLICATION NUMBER: US 60/049,018
; PRIOR FILING DATE: 1997-06-09
```

```
/ NUMBER OF SEQ ID NOS: 303
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 141
/ LENGTH: 2645
/ TYPE: DNA
/ ORGANISM: Herpes simplex
US-09-994-404-141
```

```
Query Match      68.5%; Score 22.6; DB 43; Length 2645;
Best Local Similarity 95.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 CAAACTCGTGCTCTCCAGCATG 28
      |||||:|||||:|||||:|||||:
Db      755 CAAACTCGTGCTCTCCAGCATG 777
```

```
RESULT 9
US-08-123-456-78/c
/ Sequence 78, Application US/08123456
/ GENERAL INFORMATION:
```

```
/ APPLICANT: ESSER, KLAUS M.
/ APPLICANT: CHAN, JOHN Y.
/ APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
/ APPLICANT: DELVECCHIO, ALFRED MICHAEL
/ APPLICANT: DILON, SUSAN B.
/ APPLICANT: LEARY, JEFFREY JOSEPH
/ APPLICANT: SUTTON, DAVID
/ TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
/ FILE REFERENCE: P50583
/ CURRENT APPLICATION NUMBER: US/08/123,456
/ PRIOR FILING DATE: 1997-03-03
/ PRIOR APPLICATION NUMBER: US 60/030,279
/ PRIOR FILING DATE: 1999-11-04
/ PRIOR APPLICATION NUMBER: US 60/049,018
/ PRIOR FILING DATE: 1997-06-09
/ NUMBER OF SEQ ID NOS: 303
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 78
/ LENGTH: 2841
/ TYPE: DNA
/ ORGANISM: Herpes simplex
/ NAME/KEY: unknown
/ LOCATION: (2591) (2825) (2833)
/ OTHER INFORMATION:
US-08-123-456-78
```

```
Query Match      68.5%; Score 22.6; DB 6; Length 2841;
Best Local Similarity 95.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 CAAACTCGTGCTCTCCAGCATG 28
      |||||:|||||:|||||:|||||:
Db      2071 CAAACTCGTGCTCTCCAGCATG 2049
```

```
RESULT 10
US-09-297-477A-78/c
/ Sequence 78, Application US/09297477A
/ GENERAL INFORMATION:
/ APPLICANT: ESSER, KLAUS M.
/ APPLICANT: CHAN, JOHN Y.
/ APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
/ APPLICANT: DELVECCHIO, ALFRED MICHAEL
/ APPLICANT: DILON, SUSAN B.
/ APPLICANT: LEARY, JEFFREY JOSEPH
/ APPLICANT: SUTTON, DAVID
/ TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
/ FILE REFERENCE: P50583
/ CURRENT APPLICATION NUMBER: US/09/297,477A
```

```
/ CURRENT FILING DATE: 1999-06-29
/ PRIOR APPLICATION NUMBER: US 60/030,279
/ PRIOR FILING DATE: 1999-11-04
/ PRIOR APPLICATION NUMBER: US 60/049,018
/ PRIOR FILING DATE: 1997-06-09
/ NUMBER OF SEQ ID NOS: 303
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 78
/ LENGTH: 2841
/ TYPE: DNA
/ ORGANISM: Herpes simplex
/ NAME/KEY: unsure
/ LOCATION: (2591) (2825) (2833)
/ OTHER INFORMATION: where n can equal A,G,C,T/U
US-09-297-477A-78
```

```
Query Match      68.5%; Score 22.6; DB 18; Length 2841;
Best Local Similarity 95.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 CAAACTCGTGCTCTCCAGCATG 28
      |||||:|||||:|||||:|||||:
Db      2071 CAAACTCGTGCTCTCCAGCATG 2049
```

```
RESULT 11
US-09-994-404-78/c
/ Sequence 78, Application US/09994404
/ GENERAL INFORMATION:
```

```
/ APPLICANT: ESSER, KLAUS M.
/ APPLICANT: CHAN, JOHN Y.
/ APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN
/ APPLICANT: DELVECCHIO, ALFRED MICHAEL
/ APPLICANT: DILON, SUSAN B.
/ APPLICANT: LEARY, JEFFREY JOSEPH
/ APPLICANT: SUTTON, DAVID
/ TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES
/ FILE REFERENCE: P50583
/ CURRENT APPLICATION NUMBER: US/09/994,404
/ PRIOR FILING DATE: 2001-11-26
/ PRIOR APPLICATION NUMBER: 09/297,477
/ PRIOR FILING DATE: 1999-06-29
/ PRIOR APPLICATION NUMBER: US 60/030,279
/ PRIOR FILING DATE: 1999-11-04
/ PRIOR APPLICATION NUMBER: US 60/049,018
/ PRIOR FILING DATE: 1997-06-09
/ NUMBER OF SEQ ID NOS: 303
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 78
/ LENGTH: 2841
/ TYPE: DNA
/ ORGANISM: Herpes simplex
/ NAME/KEY: unsure
/ LOCATION: (2591) (2825) (2833)
/ OTHER INFORMATION: where n can equal A,G,C,T/U
US-09-994-404-78
```

```
Query Match      68.5%; Score 22.6; DB 43; Length 2841;
Best Local Similarity 95.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      6 CAAACTCGTGCTCTCCAGCATG 28
      |||||:|||||:|||||:|||||:
Db      2071 CAAACTCGTGCTCTCCAGCATG 2049
```

```
RESULT 12
US-08-123-456-217
/ Sequence 217, Application US/08123456
/ GENERAL INFORMATION:
```

APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/08/123,456  
PRIOR FILING DATE: 1997-03-03  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 217  
LENGTH: 117213  
TYPE: DNA  
ORGANISM: Herpes simplex  
FEATURE:  
NAME/KEY: unknown  
LOCATION: (100831) (100832) (100833) (111047) (111050) (111051)  
OTHER INFORMATION:  
US-08-123-456-217

Query Match 68.5%; Score 22.6; DB 6; Length 117213;  
Best Local Similarity 95.7%; Pred. No. 2.8e+02;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CAAACTGCTGTCCTCCAGCATG 28  
Db 45257 CAAACTGCTGTCCTCCAGCATG 45279

RESULT 13  
US-09-297-477A-217  
Sequence 217, Application US/09297477A  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/09/297,477A  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 217  
LENGTH: 117213  
TYPE: DNA  
ORGANISM: Herpes simplex  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (100831) (100832) (100833) (111047) (111050) (111051)  
OTHER INFORMATION: where n can equal A,G,C,T/U  
US-09-297-477A-217

Query Match 68.5%; Score 22.6; DB 18; Length 117213;  
Best Local Similarity 95.7%; Pred. No. 2.8e+02;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CAAACTGCTGTCCTCCAGCATG 28  
Db 45257 CAAACTGCTGTCCTCCAGCATG 45279

RESULT 14  
US-09-994-404-217  
Sequence 217, Application US/09994404  
GENERAL INFORMATION:  
APPLICANT: ESSER, KLAUS M.  
APPLICANT: CHAN, JOHN Y.  
APPLICANT: DABROWSKI-AMARAL, CHRISTINE ELLEN  
APPLICANT: DELVECCHIO, ALFRED MICHAEL  
APPLICANT: DILLON, SUSAN B.  
APPLICANT: LEARY, JEFFREY JOSEPH  
APPLICANT: SUTTON, DAVID  
TITLE OF INVENTION: NOVEL CODING SEQUENCES FROM HERPES  
FILE REFERENCE: P50583  
CURRENT APPLICATION NUMBER: US/09/994,404  
PRIOR FILING DATE: 2001-11-26  
PRIOR APPLICATION NUMBER: 09/297,477  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: US 60/030,279  
PRIOR FILING DATE: 1999-11-04  
PRIOR APPLICATION NUMBER: US 60/049,018  
PRIOR FILING DATE: 1997-06-09  
NUMBER OF SEQ ID NOS: 303  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 217  
LENGTH: 117213  
TYPE: DNA  
ORGANISM: Herpes simplex  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (100831) (100832) (100833) (111047) (111050) (111051)  
OTHER INFORMATION: where n can equal A,G,C,T/U  
US-09-994-404-217

Query Match 68.5%; Score 22.6; DB 43; Length 117213;  
Best Local Similarity 95.7%; Pred. No. 2.8e+02;  
Matches 22; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 CAAACTGCTGTCCTCCAGCATG 28  
Db 45257 CAAACTGCTGTCCTCCAGCATG 45279

RESULT 15  
PCT-US01-11372-8  
Sequence 8, Application PC/TUS0111372  
GENERAL INFORMATION:  
APPLICANT: ORSON, FRANK  
APPLICANT: KINSEY, BERMA  
APPLICANT: BHOOGAL, BALBIR  
TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION  
FILE REFERENCE: P01949US1/10004014  
CURRENT APPLICATION NUMBER: PCT/US01/11372  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: 60/195,680  
PRIOR FILING DATE: 2000-04-07  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 8  
LENGTH: 154746  
TYPE: DNA  
ORGANISM: HERPESVIRUS 2  
PCT-US01-11372-8

Query Match 68.5%; Score 22.6; DB 1; Length 154746;  
Best Local Similarity 95.7%; Pred. No. 2.9e+02;

Matches	22;	Conservative	1;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	6	CAAACTCGTGTCTCTCCAGCATG	28						
Db	54077	CAAACTCGTGTCTCTCCAGCATG	54099						

Search completed: September 16, 2003, 22:44:57  
Job time : 2395.32 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 18:37:15 ; Search time 51.9505 Seconds  
(without alignments)  
535.803 Million cell updates/sec

Title: US-09-594-065-4

Perfect score: 33  
Sequence: 1 ccctgcaactcgtgkctccacgacgacgag 33

Scoring table: IDENTITY NUC  
Gapop 10'-0', Gapext 1.0

Searched: 606776 seqs, 421745864 residues

Total number of hits satisfying chosen parameters: 1213552

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_NA\_New: \*  
1: /cgn2\_6/ptodata/2/pna/PCT\_US06\_NEW\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/pna/US07\_NEW\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq: \*  
7: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20.2	61.2	2893	US-60-496-393-1	Sequence 1, Appl
2	20.2	61.2	3059	PCT-US03-21510-30	Sequence 30, Appl
3	20.2	61.2	3233	US-60-490-890-2277	Sequence 2277, Ap
4	20.2	60.6	36377	US-60-487-610-19291	Sequence 19291, A
5	20.2	60.6	46883	US-60-487-610-19887	Sequence 19887, A
6	20.2	60.6	4813087	US-09-947-914-75	Sequence 75, Appl
7	19.8	60.0	817	US-60-493-007-3124	Sequence 3124, Ap
8	19.8	60.0	904	US-60-487-610-1136	Sequence 1136, Ap
9	19.8	60.0	3679	US-09-908-576-244	Sequence 244, Ap
10	19.8	60.0	26000	US-60-495-114-16801	Sequence 16801, A
11	19.8	60.0	29102	US-60-487-610-19825	Sequence 19825, A
12	19.8	60.0	29102	US-60-485-450-12335	Sequence 12335, A
13	19.8	60.0	60476	US-60-487-610-19824	Sequence 19824, A
14	19.8	60.0	60476	US-60-485-450-12207	Sequence 12207, A
15	19.8	60.0	62312	US-60-495-114-16591	Sequence 16591, A
16	19.8	57.6	1801	US-10-425-114A-30109	Sequence 30109, A
17	19.8	57.6	513030	US-60-487-610-19737	Sequence 19737, A
18	18.8	57.0	3609	US-60-493-007-3356	Sequence 3356, Ap
19	18.8	57.0	4749	PCT-US03-073408-3	Sequence 3, Appl
20	18.8	57.0	4974	PCT-US03-073408-10	Sequence 10, Appl
21	18.8	57.0	43782	US-60-485-450-12181	Sequence 12181, A
22	18.8	57.0	47493	PCT-US02-36071A-55	Sequence 55, Appl
23	18.6	56.4	2600	US-10-425-114A-34704	Sequence 34704, A
24	18.6	56.4	83080	US-09-897-516A-423	Sequence 423, App
25	18.6	56.4	83080	US-09-897-516A-3852	Sequence 3852, Ap
26	18.6	56.4	83080	US-09-897-516A-3853	Sequence 3853, Ap

27	18.6	56.4	83080	US-09-897-516A-3854	Sequence 3854, Ap
28	18.6	56.4	83080	US-09-897-516A-3855	Sequence 3855, Ap
29	18.6	56.4	83080	US-09-897-516A-3856	Sequence 3856, Ap
30	18.6	56.4	83080	US-09-897-516A-3857	Sequence 3857, Ap
31	18.6	56.4	83080	US-09-897-516A-3858	Sequence 3858, Ap
32	18.6	56.4	83080	US-09-897-516A-3859	Sequence 3859, Ap
33	18.6	56.4	83080	US-09-897-516A-3860	Sequence 3860, Ap
34	18.6	56.4	83080	US-09-897-516A-3861	Sequence 3861, Ap
35	18.6	56.4	83080	US-09-897-516A-3862	Sequence 3862, Ap
36	18.6	56.4	83080	US-09-897-516A-3863	Sequence 3863, Ap
37	18.6	56.4	83080	US-09-897-516A-3864	Sequence 3864, Ap
38	18.6	56.4	83080	US-09-897-516A-3865	Sequence 3865, Ap
39	18.6	56.4	83080	US-09-897-516A-3866	Sequence 3866, Ap
40	18.6	56.4	83080	US-09-897-516A-3867	Sequence 3867, Ap
41	18.6	56.4	83080	US-09-897-516A-3868	Sequence 3868, Ap
42	18.6	56.4	83080	US-09-897-516A-3869	Sequence 3869, Ap
43	18.6	56.4	83080	US-09-897-516A-3870	Sequence 3870, Ap
44	18.6	56.4	83080	US-09-897-516A-3871	Sequence 3871, Ap
45	18.6	56.4	83080	US-09-897-516A-3872	Sequence 3872, Ap

## ALIGNMENTS

```
RESULT 1
US-60-496-393-1/c
; Sequence 1, Application US/60496393
; GENERAL INFORMATION:
; APPLICANT: Amit, Ido
; APPLICANT: Yakir, Liat
; TITLE OF INVENTION: POLYNUCLEOTIDES, POLYPEPTIDES AND ANTIBODIES AND USE THEREOF IN
; TITLE OF INVENTION: TREATING TSG101-ASSOCIATED DISEASES
; FILE REFERENCE: 25725
; CURRENT APPLICATION NUMBER: US/60496,393
; CURRENT FILING DATE: 2003-08-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-496-393-1

Query Match      61.2%  Score 20.2;  DB 7;  Length 2893;
Best Local Similarity 81.5%;  Pred. No. 24;
Matches 22;  Conservative 1;  Mismatches 4;  Indels 0;  Gaps 0;

Cy      1  CCCTGCAACTCGTGTCTCCACGACAT 27
Db      902 CCCTGCAACTCGTGTCTCCACGACAT 876

RESULT 2
PCT-US03-21510-30
; Sequence 30, Application PC/TUS0321510
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: NP215 AS MODIFIERS OF THE p21 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX03-046C-PC
; CURRENT APPLICATION NUMBER: PCT/US03/21510
; CURRENT FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/394,795
; PRIOR FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/401,739
; PRIOR FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: US 60/411,010
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 60/437,158
; PRIOR FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
```

LENGTH: 3059  
TYPE: DNA  
ORGANISM: Homo sapiens  
PCT-US03-21510-30

Query Match 61.2%; Score 20.2; DB 1; Length 3059;  
Best Local Similarity 75.8%; Pred. No. 24;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CCTGCAAACTCGTGTCTCCAGCATGAGG 33  
Db CCTGCAAACTCGTGTCTCCAGCATGAGG 829

RESULT 3  
US-60-490-890-2277  
Sequence 2277, Application US/60490890  
GENERAL INFORMATION:  
APPLICANT: Li, Martha  
APPLICANT: Rupnow, Brent A.  
APPLICANT: Webster, Kevin R.  
APPLICANT: Jackson, Donald  
APPLICANT: Wong, Tai W.  
TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION  
FILE REFERENCE: D0310 PSP  
CURRENT APPLICATION NUMBER: US/60/490,890  
CURRENT FILING DATE: 2003-07-29  
NUMBER OF SEQ ID NOS: 2779  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 2277  
LENGTH: 3233  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-490-890-2277

Query Match 61.2%; Score 20.2; DB 7; Length 3233;  
Best Local Similarity 75.8%; Pred. No. 24;  
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 CCTGCAAACTCGTGTCTCCAGCATGAGG 33  
Db CCTGCAAACTCGTGTCTCCAGCATGAGG 818

RESULT 4  
US-60-487-610-19291/C  
Sequence 19291, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: HUANG, Hongjin  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,  
TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17  
NUMBER OF SEQ ID NOS: 97101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19291  
LENGTH: 36377  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-487-610-19291

Query Match 60.6%; Score 20; DB 7; Length 36377;  
Best Local Similarity 76.7%; Pred. No. 43;  
Matches 23; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

2 CCTGCAAACTCGTGTCTCCAGCATGAG 31  
Db CCTGCAAACTCGTGTCTCCAGCATGAG 4982

RESULT 5  
US-60-487-610-19887  
Sequence 19887, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: HUANG, Hongjin  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,  
TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17  
NUMBER OF SEQ ID NOS: 97101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19887  
LENGTH: 46883  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-487-610-19887

Query Match 60.6%; Score 20; DB 7; Length 46883;  
Best Local Similarity 76.7%; Pred. No. 45;  
Matches 23; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

1 CCTGCAAACTCGTGTCTCCAGCATGCA 30  
Db CCTGCAAACTCGTGTCTCCAGCATGCA 14057

RESULT 6  
US-09-947-914-75/C  
Sequence 75, Application US/09947914  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig  
TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON CH  
TITLE OF INVENTION: 21, METHODS OF DETECTION, AND USES THEREOF  
FILE REFERENCE: CL001298  
CURRENT APPLICATION NUMBER: US/09/947,914  
CURRENT FILING DATE: 2001-09-07  
NUMBER OF SEQ ID NOS: 75  
SEQ ID NO 75  
LENGTH: 4813087  
TYPE: DNA  
ORGANISM: HUMAN  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(4813087)  
OTHER INFORMATION: n = A,T,C or G  
US-09-947-914-75

Query Match 60.6%; Score 20; DB 5; Length 4813087;  
Best Local Similarity 76.7%; Pred. No. 74;  
Matches 23; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

1 CCTGCAAACTCGTGTCTCCAGCATGCA 30  
Db CCTGCAAACTCGTGTCTCCAGCATGCA 3040446

RESULT 7  
US-60-493-007-3124/C  
Sequence 3124, Application US/60493007  
GENERAL INFORMATION:  
APPLICANT: Mendrick, Donna  
APPLICANT: Porter, Mark  
APPLICANT: Johnson, Kory  
APPLICANT: Higgins, Brandon  
APPLICANT: Caselle, Arthur  
APPLICANT: Elashoff, Michael  
TITLE OF INVENTION: Primary rat hepatocyte toxicity modeling  
FILE REFERENCE: 44921-5113-01-US  
CURRENT APPLICATION NUMBER: US/60/493,007  
CURRENT FILING DATE: 2003-08-07

PRIOR APPLICATION NUMBER: US 60/353,171  
PRIOR FILING DATE: 2002-02-04  
PRIOR APPLICATION NUMBER: US 60/363,534  
PRIOR FILING DATE: 2002-03-13  
PRIOR APPLICATION NUMBER: US 60/371,135  
PRIOR FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: US 60/371,134  
PRIOR FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: US 60/370,248  
PRIOR FILING DATE: 2002-04-08  
PRIOR APPLICATION NUMBER: US 60/371,150  
PRIOR FILING DATE: 2002-04-10  
PRIOR APPLICATION NUMBER: US 60/371,413  
PRIOR FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: US 60/373,601  
PRIOR FILING DATE: 2002-04-19  
PRIOR APPLICATION NUMBER: US 60/374,139  
PRIOR FILING DATE: 2002-04-22  
PRIOR APPLICATION NUMBER: US 60/394,253  
PRIOR FILING DATE: 2002-07-09  
Remaining Prior Application data removed - See file Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 3518  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3124  
LENGTH: 817  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. NM\_019194  
US-60-493-007-3124

Query Match 60.0%; Score 19.8; DB 7; Length 817;  
Best Local Similarity 84.0%; Pred. No. 29;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 3 CTGCAACTCGTGTCTCTCCAGCAT 27  
Db 107 CTGCACTCGTGTCTCTCCAGCTT 83

RESULT 8  
US-60-487-610-1136  
Sequence 1136, Application US/60487610  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: HUANG, Hongjin  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,  
FILE REFERENCE: C1001469  
CURRENT APPLICATION NUMBER: US/60/487,610  
CURRENT FILING DATE: 2003-07-17  
NUMBER OF SEQ ID NOS: 92101  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1136  
LENGTH: 904  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-487-610-1136

Query Match 60.0%; Score 19.8; DB 7; Length 904;  
Best Local Similarity 72.7%; Pred. No. 29;  
Matches 24; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Oy 1 CCTGCAAACTCGTGTCTCTCCAGTCAGGG 33  
Db 782 CCTGCACTCTCCAGTGTCTCCAGCATCTGGG 814

RESULT 9  
US-09-908-576-244/C  
Sequence 244, Application US/09908576  
GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Borstein, David  
APPLICANT: Deenoyers, Luc  
APPLICANT: Baton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Geo, Wei-Qiang  
APPLICANT: Geber, Hanspeter  
APPLICANT: Geritsen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth J.  
APPLICANT: Kijavini, Ivar J.  
APPLICANT: Mather, Dennis P.  
APPLICANT: Pan, James  
APPLICANT: Paonli, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tunas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/908,576  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US/09/665,350B  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
Remaining Prior Application data removed - See file Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 244  
LENGTH: 3679  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-09-908-576-244

Query Match 60.0%; Score 19.8; DB 5; Length 3679;  
Best Local Similarity 84.0%; Pred. No. 37;  
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 CCTGCAAACTCGTGTCTCTCCAGC 25  
Db 1297 CCTGCAAACTGTGTCTCTCCAGC 1273

RESULT 10  
US-60-495-114-16801  
Sequence 16801, Application US/60495114  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

```

; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001480
; CURRENT APPLICATION NUMBER: US/60/495,114
; CURRENT FILING DATE: 2003-08-15
; NUMBER OF SEQ ID NOS: 91238
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 16801
; LENGTH: 26000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(26000)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-60-495-114-16801

Query Match          60.0%; Score 19.8; DB 7; Length 26000;
Best Local Similarity 72.7%; Pred. No. 50;
Matches 24; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTGCAAACTCGTGTCTCTCCAGCATGACGG 33
DB 12108 CCTGCACTCTCCAGTGTTCAGCATTTCTGGG 12140

RESULT 11
US-60-487-610-19825
; Sequence 19825, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 19825
; LENGTH: 29102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-19825

Query Match          60.0%; Score 19.8; DB 7; Length 29102;
Best Local Similarity 72.7%; Pred. No. 51;
Matches 24; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTGCAAACTCGTGTCTCTCCAGCATGACGG 33
DB 12108 CCTGCACTCTCCAGTGTTCAGCATTTCTGGG 12140

RESULT 12
US-60-485-450-12335
; Sequence 12335, Application US/60485450
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/485,450
; CURRENT FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 12335
; LENGTH: 29102
; TYPE: DNA
US-60-495-114-16591/c
```

```

; ORGANISM: Homo sapiens
US-60-485-450-12335

Query Match          60.0%; Score 19.8; DB 7; Length 29102;
Best Local Similarity 72.7%; Pred. No. 51;
Matches 24; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTGCAAACTCGTGTCTCTCCAGCATGACGG 33
DB 12108 CCTGCACTCTCCAGTGTTCAGCATTTCTGGG 12140

RESULT 13
US-60-487-610-19824/c
; Sequence 19824, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; FILE REFERENCE: CL001469
; CURRENT APPLICATION NUMBER: US/60/487,610
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 19824
; LENGTH: 60476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-487-610-19824

Query Match          60.0%; Score 19.8; DB 7; Length 60476;
Best Local Similarity 72.7%; Pred. No. 58;
Matches 24; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTGCAAACTCGTGTCTCTCCAGCATGACGG 33
DB 58060 CCTGCACTCTCCAGTGTTCAGCATTTCTGGG 58028

RESULT 14
US-60-485-450-12207/c
; Sequence 12207, Application US/60485450
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; FILE REFERENCE: CL001470
; CURRENT APPLICATION NUMBER: US/60/485,450
; CURRENT FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO: 12207
; LENGTH: 60476
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-485-450-12207

Query Match          60.0%; Score 19.8; DB 7; Length 60476;
Best Local Similarity 72.7%; Pred. No. 58;
Matches 24; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 CCTGCAAACTCGTGTCTCTCCAGCATGACGG 33
DB 58060 CCTGCACTCTCCAGTGTTCAGCATTTCTGGG 58028

RESULT 15
US-60-495-114-16591/c
```

```

: Sequence 16591, Application US/60495114
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
: TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
: TITLE OF INVENTION: USES THEREOF
: FILE REFERENCE: CL001480
: CURRENT APPLICATION NUMBER: US/60/495,114
: CURRENT FILING DATE: 2003-08-15
: NUMBER OF SEQ ID NOS: 91238
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 16591
: LENGTH: 62312
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (1)..(62312)
: OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1
US-60-495-114-16591

```

Query Match	60.0%	Score 19.8;	DB 7;	Length 62312;
Best Local Similarity	72.7%;	Pred. No. 58;		
Matches 24;	Conservative 1;	Mismatches 8;	Indels 0;	Gaps 0;

QY 1 CCGTCGAAACTCGTGTCTCTCAGCATGACGGG 33  
| | | | | : | | | | | | | | | | | |  
Db 58061 CCGTCGACTCTCCAGTTGTTCCAGCANTTGGG 58029

Search completed: September 16, 2003, 22:48:21  
Job time : 59.9505 Secs

**BEST AVAILABLE COPY**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:35:29 ; Search time 394.752 Seconds  
(without alignments)  
1865.405 Million cell updates/sec

Title: US-09-594-065-2

Perfect score: 18  
Sequence: 1 agcagccgcgcgcctcctg 18

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBankl:  
1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_srs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_srs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pla:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	18	100.0	798	6 103657	103657 Sequence 5
C 2	18	100.0	1478	14 S74390	S74390 gBpath/U127
C 3	18	100.0	1488	14 S65444	S65444 U127-glycop
C 4	18	100.0	1857	14 HS1GSM	K02720 HSV1 (mutan
C 5	18	100.0	2088	6 E03093	E03093 DNA encodin
C 6	18	100.0	2586	6 E00358	E00358 DNA encodin
C 7	18	100.0	2712	6 BD145144	BD145144 Herpes a1
C 8	18	100.0	2712	6 BD165793	BD165793 Vaccine f
C 9	18	100.0	2712	6 BD165795	BD165795 Vaccine f
C 10	18	100.0	2715	6 BD145143	BD145143 Herpes a1
C 11	18	100.0	2715	14 AF097023	AF097023 Human her
C 12	18	100.0	2715	14 AF295528	AF295528 Human her
C 13	18	100.0	2715	14 AF311740	AF311740 Human her
C 14	18	100.0	2715	14 HHU49121	U49121 Human herpe
C 15	18	100.0	2809	14 AF259899	AF259899 Human her
C 16	18	100.0	2817	14 BHU14662	BHU14662 Baboon herp
C 17	18	100.0	2818	14 AF021340	AF021340 Human her
C 18	18	100.0	2846	14 HHU12172	U12172 Human herpe
C 19	18	100.0	2846	14 HHU12174	U12174 Human herpe
C 20	18	100.0	2855	14 HHU12173	U12173 Human herpe
C 21	18	100.0	2855	14 HHU12175	U12175 Human herpe
C 22	18	100.0	2925	6 E03115	E03115 DNA encodin
C 23	18	100.0	2943	6 117281	M57388 Simian herp
C 24	18	100.0	2943	14 H5VGB0	U14663 Baboon herp
C 25	18	100.0	2973	6 E00883	E00883 Fragment of
C 26	18	100.0	3098	6 E00883	E00883 Fragment of
C 27	18	100.0	3324	14 HS2GB	M1923 HSV2 glycop
C 28	18	100.0	3461	6 E01195	E01195 DNA sequenc
C 29	18	100.0	3465	6 E03024	E03024 DNA encodin
C 30	18	100.0	3465	6 E03092	E03092 DNA encodin
C 31	18	100.0	3465	6 E03112	E03112 DNA encodin
C 32	18	100.0	3472	6 BD165792	BD165792 Vaccine f
C 33	18	100.0	3472	6 BD165794	BD165794 Vaccine f
C 34	18	100.0	3472	6 108685	I08685 Sequence 1
C 35	18	100.0	3472	6 108686	I08686 Sequence 2
C 36	18	100.0	3472	14 HS2GB3	M15118 HSV-2 (333)
C 37	18	100.0	3640	6 108361	I08361 Sequence 1
C 38	18	100.0	3643	14 HS1GBP	K03541 HSV-1 (Palc
C 39	18	100.0	3715	14 HS2GB2A	M24771 Herpes simp
C 40	18	100.0	3755	6 E03025	E03025 DNA encodin
C 41	18	100.0	3755	6 E03113	E03113 DNA encodin
C 42	18	100.0	3758	14 HS1GB	K01760 HSV1 (KOS)
C 43	18	100.0	3996	6 E03026	E03026 DNA encodin
C 44	18	100.0	3996	6 E03114	E03114 DNA encodin
C 45	18	100.0	3997	6 E00357	E00357 DNA encodin

## ALIGNMENTS

RESULT 1  
103657/c  
LOCUS 103657  
DEFINITION Sequence 5 from Patent US 4642333.  
ACCESSION 103657  
VERSION 103657.1 GI:268613  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 798)  
AUTHORS Person,S.  
TITLE Immunologically reactive non-glycosylated amino acid chains of glycoprotein B of herpes virus types 1 and 2  
JOURNAL Patent: US 4642333-A 5 10-FEB-1987;

600 Locust La.; State College, PA  
**FEATURES**  
 Location/Qualifiers  
 source 1..798 /organism="unknown"  
 BASE COUNT 155 a 261 c 254 g 128 t  
 ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 798;  
 Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCCGCTGCTCTTG 18  
 |||||  
 182 AGCAGCCGCTGCTCTTG 165

RESULT 2  
 S74390 1478 bp DNA linear VRL 30-APR-1995  
 LOCUS GBpath/UL27=glycoprotein B [herpes simplex virus type 1 HSV-1, ANG  
 DEFINITION GBpath, Genomic Mutant, 1478 nt].  
 S74390  
 VERSION S74390.1 GI:786563  
 KEYWORDS Human herpesvirus 1  
 SOURCE Human herpesvirus 1  
 ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Simplexvirus.

REFERENCE 1 (bases 1 to 1478)  
 Kozak, M., Bacik, J., Rajcani, J. and Kaerner, H.C.  
 Replacement of glycoprotein B gene in the herpes simplex virus type 1 strain ANGpath DNA by that originating from nonpathogenic strain KOS reduces the pathogenicity of recombinant virus  
 Acta Virol. 38 (2), 77-88 (1994)

JOURNAL MEDLINE 95067449  
 PUBMED 7976866  
 REMARK Genbank staff at the National Library of Medicine created this entry [NCBI gisbseq 158378] from the original journal article.  
 This sequence comes from Fig. 8.  
 Location/Qualifiers

FEATURES  
 source 1..1478 /organism="Human herpesvirus 1"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10298"  
 1..1478 /partial  
 /gene="gbpath/UL27"  
 /note="glycoprotein B"  
 BASE COUNT 236 a 451 c 528 g 263 t  
 ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 1478;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCCGCTGCTCTTG 18  
 |||||  
 348 AGCAGCCGCTGCTCTTG 365

RESULT 3  
 S65444 1488 bp DNA linear VRL 21-APR-2003  
 LOCUS UL27=glycoprotein B [herpes simplex virus type 1 HSV-1, ANG,  
 DEFINITION Genomic, 1488 nt].  
 S65444  
 VERSION S65444.1 GI:415494  
 KEYWORDS Human herpesvirus 1  
 SOURCE Human herpesvirus 1  
 ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Simplexvirus.

REFERENCE 1 (bases 1 to 1488)

Yuhaz, S.A. and Stevens, J.G.  
**AUTHORS**  
 TITLE Glycoprotein B is a specific determinant of herpes simplex virus type 1 neuroinvasiveness  
 JOURNAL J Virol. 67 (10), 5948-5954 (1993)  
 MEDLINE 8396662  
 PUBMED 8396662  
 REMARK Genbank staff at the National Library of Medicine created this entry [NCBI gisbseq 137392] from the original journal article.  
 This sequence comes from Fig. 7.  
 Location/Qualifiers

FEATURES  
 source 1..1488 /organism="Human herpesvirus 1"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10298"  
 <1..>1488 /gene="UL27"  
 <1..>1488 /gene="UL27"  
 /note="This sequence comes from Fig. 7"  
 /codon\_start=1 /product="glycoprotein B"  
 /protein\_id="AAP13961.1"  
 /db\_xref="GI:30027723"  
 /translation="VDSFYARDLTTKADATPTTNLTTPKFTVAMDMVPRPSVCT /translating="VDGSEFRSSDAISTFTNTLVEPLSRVLDGDCIGKARDMDR ITPARYNTHIRVQPOYVLANGLIAYOPPLSTLAEIYREHLREGSRKPPPTP PPGSANSVVERIKTSSIEPARLOFTYNIQRVNDMLGVALAMCELOHETLM NEARKLNATASATVGRVRSATGVAATCVPAADVIVONSMSISRPACV SRPLVFRYEDGQPLVEGOLGSENNELRLTRDIECTYGHRRYFTFGGYVPRAY SHQLSRADITTVSTFDLNTIMLEDEHEVPELVYRHEIKDGLDYDEVORNDLHD LRFADIDVTHADANAMFAGAFEGMGDGRVAVGVKMGIVGVVSAVSGVSSFM SNPFALNVLVLVAGLAAAPFAPFVVRLOSPMKALYPLTTKEU"

BASE COUNT 260 a 530 c 459 g 239 t  
 ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 1488;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCCGCTGCTCTTG 18  
 |||||  
 1136 AGCAGCCGCTGCTCTTG 1119

RESULT 4  
 HS1GBM/c 1857 bp DNA linear VRL 02-AUG-1993  
 LOCUS HS1GBM (mutant strain ts85), glycoprotein B (gB) gene.  
 DEFINITION K02720  
 ACCESSION K02720.1 GI:330087  
 VERSION glycoprotein.  
 KEYWORDS Human herpesvirus 1  
 SOURCE Human herpesvirus 1  
 ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Simplexvirus.

REFERENCE 1 (bases 1 to 1857)  
 Bzik, D.J., Fox, B.A., Deluca, N.A. and Person, S.  
 Nucleotide sequence of a region of the herpes simplex virus type 1 gB glycoprotein gene: mutations affecting rate of virus entry and cell fusion  
 Virology 137 (1), 185-190 (1984)

JOURNAL MEDLINE 84301857  
 PUBMED 6089415  
 COMMENT Original source text: HSV1 strain ts85 DNA, clone pTBG-B51.  
 HSV1 strain ts85 is a mutant of strain HEM chat is temperature dependent for virus growth, production of gB, and fusion of Vero and HEL cells. The syn and ts lesions of ts85 are separable by recombination, the syn defect being located between 0.345 and 0.355 map units, and the ts defect being located between 0.360 and 0.368 map units. The rate of entry determinant of ts85 is genetically separable from both the syn and ts lesions and is located between 0.350 and 0.360 map units, which is entirely within the glycoprotein B gene (0.348 to 0.366 map units).



Strain ts85 has a total of 12 base substitutions within the coding region for gb when compared to strain KOS. Of these, only two cause amino acid substitutions; a 't' to 'c' transition at bp 695 causes a Val to Ala substitution, and a 'g' to 'a' transition at bp 1610 causes an Arg to His substitution. Because the rate-of-entry determinant has been mapped to between 0.350 and 0.360 map units, the transition at bp 695 must be responsible for the difference in rate-of-entry between ts85 and KOS.

FEATURES  
source

1..1857  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10298"  
<1..1752  
/note="glycoprotein B"  
/codon\_start=1  
/protein\_id="AA045777.1"  
/db\_xref="GI:330088"

## CDS

/translation="VDSFYARDLTTKARATPTNNLTTPKFTVAMWVPRKPSVCT  
MTKQEVDEMIRSEYSGSFRSSDAISTFTTNLTVEPLSRVLDGDCIGKADAMDR  
IFARRYNATHIKVQPOYYLANGFLIAYOPLNTLAEIYRHLRQSRKPPPTP  
PPGASANASVERIKTSSIEPARLQFTYNNHQRHVMIGRAIAMGELONHETLM  
NEARKLNPAIASATGRRVSRVAMIGDYMAYSTCPVADNYIVNSMRISBRACY  
SRPLVSFRYEDQPLVBQGLGNNELRLTRDILIEPCTGHRRIYFTFGGIYFEETAY  
SHQSRADITTVSTFIDINTLIEDHEFVPLEVYTRHBIKOSGLDYTEVORNLGD  
LRFADIDIVIHADANAAMFAGIAGFEFGDIDGRAVGVVMGIVGVVSAVSGSFM  
SNPFGALVGLIVLAGLAFAFRYVRIGLQSNPMKALYPLTKELKQPTPDGSGEG  
EEGGFDEAKLAEREIRMYALVSAHTEHKAKKGTSLSKATVDMWRKRRT  
MYTQVPMKDGADBDLD"

BASE COUNT 358 a 624 c 576 g 299 t  
ORIGIN 1 bp upstream of SalI site; map coordinate 0.360.

Query Match 100.0%; Score 18; DB 14; Length 1857;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGCGCGCTGCTCTTG 18  
Db 1136 AGCAGCGCGCTGCTCTTG 1119

RESULT 5  
E03093 2088 bp DNA linear PAT 29-SEP-1997  
LOCUS DNA encoding truncated gb of herpes simplex virus type1 (HSV-1).  
DEFINITION E03093  
ACCESSION E03093  
VERSION E03093.1 GI:2171311  
KEYWORDS JP 1991218397-A/2.  
SOURCE Human herpesvirus 1  
ORGANISM Human herpesvirus 1  
Virusess; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
1 (bases 1 to 2088)  
Fujisawa,Y., Hinuma,K., Asakawa,N. and Otake,S.  
POLYPEPTIDE AND ITS PRODUCTION  
Patent: JP 1991218397-A 2 25-SEP-1991;  
TAKEDA CHEM IND LTD  
OS Herpes simplex virus type1 (HSV-1)  
PN JP 1991218397-A/2

REFERENCE  
AUTHORS Fujisawa,Y., Hinuma,K., Asakawa,N. and Otake,S.  
TITLE POLYPEPTIDE AND ITS PRODUCTION  
JOURNAL Patent: JP 1991218397-A 2 25-SEP-1991;  
COMMENT  
OS Herpes simplex virus type1 (HSV-1)  
PN JP 1991218397-A/2  
PD 25-SEP-1991  
PF 21-JUN-1990 JP 1990161448  
PR 22-JUN-1989 JP 89P 158238, 30-NOV-1989 JP 89P 308941 PI  
FUJISAWA YUKIO, HINUMA KUNIKI, ASAKAWA NAOKI, OTAKA SACHIKO PC  
C07K13/00, C12N1/19, C12N15/38, C12P21/02//A61K39/245, (C12N1/19, PC  
C12R1:865),  
PC (C12P21/02, C12R1:865);  
CC strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: strain=Hiyama;  
CC \*source: clone=PHS106 delta Tth;  
FH Key Location/Qualifiers

EH  
FT CDS 1..2085  
FT /product='truncated gb of HSV-1' FT 3'UTR  
FT 2086..2088  
FT mat\_peptide 1..2082  
FT /product='truncated gb of HSV-1'.  
FT Location/Qualifiers  
1..2088  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10298"

FEATURES  
source

1..2088  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10298"

BASE COUNT 417 a 751 c 620 g 300 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 2088;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGCGCGCTGCTCTTG 18  
Db 2012 AGCAGCGCGCTGCTCTTG 1995

RESULT 6  
E00358 2586 bp DNA linear PAT 29-SEP-1997  
LOCUS DNA encoding glycoprotein B of herpes simplex virus type 2.  
DEFINITION E00358  
ACCESSION E00358.1 GI:2168645  
KEYWORDS JP 1985115529-A/2.  
SOURCE unidentified  
ORGANISM unidentified  
unclassified.  
1 (bases 1 to 2586)  
REFERENCE  
AUTHORS Sutanree,P.  
TITLE NON-GLYCOSYLATED AMINO ACID CHAIN IMMUNOLOGICALLY REACTIVE TO  
GLYCOPROTEIN OF HERPES VIRUS 1 AND 2  
Patent: JP 1985115529-A 2 22-JUN-1985;  
JOURNAL SUTANREE PASON  
COMMENT  
OS herpes simplex virus type 2  
PN JP 1985115529-A/2  
PD 22-JUN-1985  
PF 23-JUN-1984 JP 1984129915  
PR 23-JUN-1983 US 83 506986, 16-SEP-1983 US 83 532996 PI  
SUTANREE PASON  
PC A61K39/245, C07H21/04, C07K15/14, C12N15/00//C12P21/02; CC  
strandedness: Double;  
CC topology: Linear;  
CC hypothetical: No;  
CC anti-sense: No;  
CC \*source: strain=HG 52;  
CC \*source: clone=p52BX;  
CC \*source: map\_position=0.386-0.348;  
FH Key Location/Qualifiers  
FT  
FT CDS 1..2727  
FT /product='glycoprotein B of herpes simplex FT  
FT virus type 2'.  
FT Location/Qualifiers  
1..2586  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

FEATURES  
source  
1..2586  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 481 a 886 c 841 g 378 t  
ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 2586;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCAGCGCGCTGCTCTTG 18  
Db 2025 AGCAGCGCGCTGCTCTTG 2008

RESULT 7  
BD145144/c  
LOCUS BD145144 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Herpes simplex virus VP16 vaccine.  
ACCESSION BD145144  
VERSION BD145144.1 GI:27850902  
KEYWORDS JP 2002136297-A/4.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 2712)  
AUTHORS Burke, R.L. and Sekulovich, R.E.  
TITLE Herpes simplex virus VP16 vaccine  
JOURNAL Patent: JP 2002136297-A 4 14-MAY-2002;  
CHIRON CORP  
COMMENT OS HSV gB1  
PN JP 2002136297-A/4  
PD 14-MAY-2002  
PF 10-SEP-2001 JP 2001274335  
PR 02-AUG-1990 US 561,528  
PI RAE LYNN BURKE, ROSE E SEKULOVICH  
PC C12N15/09, A61K38/00, A61K39/245, A61P31/12, C07K14/03, C12N15/00,  
PC A61K37/02  
CC Herpes simplex virus VP16 vaccine  
FH Key Location/Qualifiers  
FT CDS (1)..(2712).

FEATURES  
source  
1..2712  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 513 a 949 c 853 g 397 t

ORIGIN  
Query Match 100.0%; Score 18; DB 6; Length 2712;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGGCCGCTGTCCTTG 18  
|||||  
DB 2096 AGCAGGCCGCTGTCCTTG 2079

RESULT 8  
BD165793/c  
LOCUS BD165793 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Vaccine for use in the therapeutic treatment of HSV.  
ACCESSION BD165793  
VERSION BD165793.1 GI:27871605  
KEYWORDS JP 2002167398-A/2.  
SOURCE Herpes simplex virus 7  
ORGANISM Herpes simplex virus 7  
REFERENCE 1 (bases 1 to 2712)  
AUTHORS Burke, R.L., Pachl, C. and Valenzuela, P.D.T.  
TITLE Vaccine for use in the therapeutic treatment of HSV  
JOURNAL Patent: JP 2002167398-A 2 11-JUN-2002;  
CHIRON CORP  
COMMENT OS Herpes simplex virus 7  
PN JP 2002167398-A/2  
PD 11-JUN-2002  
PF 12-SEP-2001 JP 2001277340  
PR 20-OCT-1986 US 921,213  
PI RAE LYNN BURKE, CAROL PACHL, PABLO D T VALENZUELA  
PC C07K14/035, A61K39/245, A61P31/22, C12N15/09, C12N15/00, C12N15/00,  
CC Vaccine for use in the therapeutic treatment of HSV  
FH Key Location/Qualifiers  
FT CDS (1)..(2712).

FEATURES  
source  
1..2712  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

source 1..2712  
/organism="Herpes simplex virus 7"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:57277"

BASE COUNT 490 a 938 c 884 g 400 t

ORIGIN  
Query Match 100.0%; Score 18; DB 6; Length 2712;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGGCCGCTGTCCTTG 18  
|||||  
DB 2090 AGCAGGCCGCTGTCCTTG 2073

RESULT 9  
BD165795/c  
LOCUS BD165795 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Vaccine for use in the therapeutic treatment of HSV.  
ACCESSION BD165795  
VERSION BD165795.1 GI:27871607  
KEYWORDS JP 2002167398-A/4.  
SOURCE Herpes simplex virus 7  
ORGANISM Herpes simplex virus 7  
REFERENCE 1 (bases 1 to 2712)  
AUTHORS Burke, R.L., Pachl, C. and Valenzuela, P.D.T.  
TITLE Vaccine for use in the therapeutic treatment of HSV  
JOURNAL Patent: JP 2002167398-A 4 11-JUN-2002;  
CHIRON CORP  
COMMENT OS Herpes simplex virus 7  
PN JP 2002167398-A/4  
PD 11-JUN-2002  
PF 12-SEP-2001 JP 2001277340  
PR 20-OCT-1986 US 921,213  
PI RAE LYNN BURKE, CAROL PACHL, PABLO D T VALENZUELA  
PC C07K14/035, A61K39/245, A61P31/22, C12N15/09, C12N15/00, C12N15/00,  
CC Vaccine for use in the therapeutic treatment of HSV  
FH Key Location/Qualifiers  
FT CDS (1)..(2712).

FEATURES  
source  
1..2712  
/organism="Herpes simplex virus 7"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:57277"

BASE COUNT 514 a 949 c 853 g 396 t

ORIGIN  
Query Match 100.0%; Score 18; DB 6; Length 2712;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGGCCGCTGTCCTTG 18  
|||||  
DB 2099 AGCAGGCCGCTGTCCTTG 2082

RESULT 10  
BD145143/c  
LOCUS BD145143 2715 bp DNA linear PAT 17-JAN-2003  
DEFINITION Herpes simplex virus VP16 vaccine.  
ACCESSION BD145143  
VERSION BD145143.1 GI:27850901  
KEYWORDS JP 2002136297-A/3.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 2715)  
AUTHORS Burke, R.L. and Sekulovich, R.E.  
TITLE Herpes simplex virus VP16 vaccine

JOURNAL Patent: JP 2002136297-A 3 14-MAY-2002;  
CHIRON CORP

COMMENT OS HSV GB2  
PN JP 2002136297-A/3  
PD 14-MAY-2002  
PF 10-SEP-2001 JP 2001274335  
PR 02-AUG-1990 US 561,528  
PI RAE LYNN BURKE, ROSE E SEKULOVICH  
PC C12N15/09, A61K38/00, A61K39/245, A61P31/12, C07K14/03, C12N15/00,  
A61K37/02  
PC Herpes simplex virus VP16 vaccine  
CC Key location/Qualifiers  
FT CDS (1)..(2715).

FEATURES  
source location/Qualifiers  
1..2715  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 491 a 938 c 886 g 400 t

ORIGIN

Query Match 100.0%; Score 18; DB 6; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCAGGCCGCTGTCTTG 18  
|||||  
Db 2090 ACCAGGCCGCTGTCTTG 2073

RESULT 11  
AF097023/c 2715 bp DNA linear VRL 07-APR-2000  
LOCUS Human herpesvirus 1 strain HSZP glycoprotein B (UL27) gene,  
DEFINITION complete cds.  
ACCESSION AF097023.1 GI:6165611  
VERSION AF097023  
KEYWORDS Human herpesvirus 1  
SOURCE Human herpesvirus 1  
ORGANISM Viruses; dDNA viruses, no RNA stage; Herpesviridae;  
REFERENCE 1 (bases 1 to 2715)  
AUTHORS Kosovsky, J., Vojvodova, A., Oravcova, I., Kudelova, M., Matls, J. and Rajcani, J.  
TITLE Herpes simplex virus 1 (HSV-1) strain HSZP glycoprotein B gene: comparison of mutations among strains differing in virulence  
JOURNAL Virus Genes 20 (1), 27-33 (2000)  
MEDLINE 20227371  
PUBMED 10766304  
REFERENCE 2 (bases 1 to 2715)  
AUTHORS Kosovsky, J., Vojvodova, A., Oravcova, I., Majorova, J., Kudelova, M. and Rajcani, J.  
TITLE Direct Submission  
JOURNAL Submitted (02-OCT-1998) Institute of Virology, Slovak Academy of Sciences, Dubravska cesta 9, Bratislava 842 46, Slovakia  
FEATURES  
source location/Qualifiers  
1..2715  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="HSZP"  
/db\_xref="taxon:10298"  
1..2715  
/gene="UL27"  
1..2715  
/gene="UL27"  
/function="essential for fusion between virion envelope and cell membrane"  
/note="envelope glycoprotein, gB"  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AA04615.1"  
/db\_xref="GI:6165612"

/translation="NRQAPARGCMFVVMALLGLTLGVASAPSSPFGTCVAAT  
QAANGPATPAPALGAAPTGDPKKXKKKPKNPPRPAGDNATVAAGHATLREHL  
DKASTDANFVCPPTGATVQFQEPKRCPTREGQVYTGIAVFEKNIAPVFK  
ATMYKQDVTSQVWFGHRSQPMGI FEBRAPVPEFVIDIKINAKVCSHTAYVNV  
ETTAAPHRDHETDMELKXANATRTSRCHMTTDLKYNBSRVARFRCTVACIEN  
DARSYTPDEFIATGDFVYNSPFYGTREGSHTEHTSTAADFQKQDGYADLTITKA  
RATAPPTNLTTPKFTVAMDVPVPRPCVCTMKQEDMLRSYGSFRSSDAIS  
TFTTNLTVEPLSRVLDGCIQKARAMDRI FARVYATHIKVQPOYYLANGFLI  
AYQPLSNTLAEIYVREHLREQSRKPPPTPPQASANAVERIKTSSIEFARLOF  
TYNHIORVNDMLGVAIALWCELOHHELTLMNEARKLNALASATVGRVSARMIGD  
VMAVSTCVPAADVNTVONSMRISSRPACVSRPLVSPRYEGGPIVREGOLGENNEIR  
LTRDAIECTVGRRIYFTFGGCTVYFEIATYSHQSLRDITTVTSFTIDNTIMLEDE  
FVPLEVYTRHEIKOSGLDYTBVQRNLDHREFADITDVIHADNAAAFAGLAFBE  
GMGDLGRAVGVKVMGIVGVVSAVGSFMSNPFGALVAGLIVLAGIAAFAFRVY  
MLOSNPMKALPVLTKELKNPTNDADGEGCEGDFDEAKLAEAREMIRYVALVSAM  
EHTEHAKKKGTSAISAKVTDVWVRKRNRTYTVPKMGDADDDL"

BASE COUNT 516 a 943 c 856 g 400 t

ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACCAGGCCGCTGTCTTG 18  
|||||  
Db 2099 ACCAGGCCGCTGTCTTG 2082

RESULT 12  
AF295528/c 2715 bp mRNA linear VRL 01-SEP-2001  
LOCUS Human herpesvirus 2 glycoprotein B2 (GB2) mRNA, complete cds.  
DEFINITION AF295528  
ACCESSION AF295528.1 GI:15426575  
VERSION AF295528  
KEYWORDS Human herpesvirus 2  
SOURCE Human herpesvirus 2  
ORGANISM Viruses; dDNA viruses, no RNA stage; Herpesviridae;  
REFERENCE 1 (bases 1 to 2715)  
AUTHORS Lee, H.H.  
TITLE Sequencing and high level expression of the HSV-2 strain G glycoprotein B gene using a baculovirus HcNPV  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 2715)  
TITLE Lee, H.H.  
JOURNAL Direct Submission  
AUTHORS Submitted (12-AUG-2000) Department of Biology, Konkuk University, 1 Hwayang-dong, Kwangjin-gu, Seoul 143-701, Korea  
FEATURES  
source location/Qualifiers  
1..2715  
/organism="Human herpesvirus 2"  
/mol\_type="mRNA"  
/strain="G"  
/db\_xref="taxon:10310"  
1..2715  
/gene="GB2"  
1..2715  
/gene="GB2"  
/codon\_start=1  
/product="glycoprotein B2"  
/protein\_id="AAK97852.1"  
/db\_xref="GI:15426575"  
/translation="NRGGGLICALVYGLVAASAAAPAAPASGVAATVAANG  
GPASRPVPSPTATKARKRKPKPEPEPATPPDANAATVAAGATLRAHLREIVE  
NADQFVCPPTGATVQFQEPKRCPTREGQVYTGIAVFEKNIAPVFKATMYK  
QDVTSQVWFGHRSQPMGI FEBRAPVPEFVIDIKINAKVCSHTAYVNVETTA  
PHRDHETDMELKXANATRTSRCHMTTDLKYNBSRVARFRCTVACIEN  
DARSYTPDEFIATGDFVYNSPFYGTREGSHTEHTSTAADFQKQDGYADLTITKA  
RATAPPTNLTTPKFTVAMDVPVPRPCVCTMKQEDMLRSYGSFRSSDAIS  
TFTTNLTVEPLSRVLDGCIQKARAMDRI FARVYATHIKVQPOYYLANGFLI  
AYQPLSNTLAEIYVREHLREQSRKPPPTPPQASANAVERIKTSSIEFARLOF  
TYNHIORVNDMLGVAIALWCELOHHELTLMNEARKLNALASATVGRVSARMIGD  
VMAVSTCVPAADVNTVONSMRISSRPACVSRPLVSPRYEGGPIVREGOLGENNEIR  
LTRDAIECTVGRRIYFTFGGCTVYFEIATYSHQSLRDITTVTSFTIDNTIMLEDE  
FVPLEVYTRHEIKOSGLDYTBVQRNLDHREFADITDVIHADNAAAFAGLAFBE  
GMGDLGRAVGVKVMGIVGVVSAVGSFMSNPFGALVAGLIVLAGIAAFAFRVY  
MLOSNPMKALPVLTKELKNPTNDADGEGCEGDFDEAKLAEAREMIRYVALVSAM  
EHTEHAKKKGTSAISAKVTDVWVRKRNRTYTVPKMGDADDDL"

VSTCVPAADNVIVONSMRISRPACYSRPLVSFRYEDQGLIBCOLDENNELRLTR  
DALEPCTVGRHRRYFTFGGIVYFEERYAHSQSRADVTTSFIDLTIMLDEHBLP  
LAVYTRHEIKDSGLDYTEVORRNQJHDFADIDIVIHADANAAMFAGLAFEGMG  
DGRVAVKVMGVGVSAVGSFMSNPGALVGLVLAGLVAAPFAFRVYIOL  
QNRPMKALYPLTTKEIKTSDPRGVBGEGBEGAPDCAKLABAREMIRYALVLSAM  
ERTREHAKKKSALLSKVTDMWRKRNTNYTVOPNKGADDEDL"

## BASE COUNT

492 a 939 c 886 g 398 t

Query Match 100.0%; Score 18; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAGAGCCCGCTGTCTTG 18  
|||||  
Db 2099 ACAGAGCCCGCTGTCTTG 2073

RESULT 13  
AF311740/c 2715 bp DNA linear VRL 24-JAN-2001  
LOCUS Human herpesvirus 1 strain KOS glycoprotein B gene, complete cde.  
DEFINITION AF311740  
VERSION AF311740.1 GI:11321322  
KEYWORDS  
SOURCE Human herpesvirus 1  
ORGANISM Human herpesvirus 1  
Virusess; daDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirinae.

REFERENCE 1 (bases 1 to 2715)  
Percel, P.B., Frideberg, A., Parish, M.L. and Spear, P.G.  
AUTHORS Cell fusion induced by herpes simplex virus glycoproteins gB, gD,  
TITLE and gH-gL requires a gD receptor but not necessarily heparan  
JOURNAL sulfatase  
Virology 279 (1), 313-324 (2001)

JOURNAL 21066717  
PUBMED 1145912  
REFERENCE 2 (bases 1 to 2715)  
Percel, P.B., Frideberg, A., Parish, M.L. and Spear, P.G.  
AUTHORS Direct Submission  
TITLE Submitted (05-OCT-2000) Microbiology-Immunology, Northwestern  
JOURNAL University Medical School, 320 E. Superior Ave., Ward 6-241,  
Chicago, IL 60611, USA  
Location/Qualifiers

## FEATURES

source

1..2715  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="KOS"  
/db\_xref="taxon:10298"  
1..2715  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AA634116.1"  
/db\_xref="GI:11321323"  
/translation="MQGAPARGRMPVVMALLGLTGLVLSAAPSPTGVAAT  
QANGPAPAPAPGAPPTGDPKPKKKPKPPRPRAGDNATVAAGHATLREHLR  
DIAKENTDANFVCPPTGATVQFQPRRCPTPRPGASANAVERIKTSSIEPARLOF  
ATMYKYDVTASQVWFGHRSOFGMIFEDAPVPEVVDIKINAKGCRSTAKYVRNL  
ETTAHRDDEHETDMEKAPNAAATRTSRGHTTDLTKNPRVRAFRHGTTCVACIEV  
DASVYPIRNEFLATGDFVYNSMRISRPACYSRPLVSFRYEDQGLIBCOLDENNELR  
LRTDAIEPCTVGRHRRYFTFGGIVYFEERYAHSQSRADITTVSTFIDLTIMLDEH  
PVPLEVYTRHEIKDSGLDYTEVORRNQJHDFADIDIVIHADANAAMFAGLAFEGMG  
DGRVAVKVMGVGVSAVGSFMSNPGALVGLVLAGLVAAPFAFRVYIOL  
QNRPMKALYPLTTKEIKTSDPRGVBGEGBEGAPDCAKLABAREMIRYALVLSAM  
ERTREHAKKKSALLSKVTDMWRKRNTNYTVOPNKGADDEDL"

## CDS

BASE COUNT 521 a 948 c 849 g 397 t  
ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAGAGCCCGCTGTCTTG 18  
|||||  
Db 2099 ACAGAGCCCGCTGTCTTG 2082

RESULT 14  
HHU49121/c 2715 bp DNA linear VRL 08-MAR-1996  
LOCUS Human herpesvirus 1 strain ANG patch, glycoprotein B (gB) gene,  
DEFINITION HHU49121  
VERSION HHU49121  
KEYWORDS complete cds.  
SOURCE Human herpesvirus 1  
ORGANISM Human herpesvirus 1  
Virusess; daDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirinae.

REFERENCE 1 (bases 1 to 2715)  
Holland, T.C. and Saharikhiz-Langrood, A.  
AUTHORS Direct Submission  
TITLE Submitted (13-FEB-1996) Dept. of Immunology and Microbiology, Wayne  
JOURNAL State University, 540 E. Canfield Ave., Detroit, MI 48201, USA  
Location/Qualifiers

## FEATURES

source

1..2715  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="ANG patch"  
/db\_xref="taxon:10298"  
1..2715  
/gene="gB"  
/note="HSV-1 UL27 open reading frame"  
1..2715  
/gene="gB"  
/note="envelope glycoprotein"  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AA491805.1"  
/db\_xref="GI:1218050"

## CDS

translation="MQGAPARGRMPVVMALLGLTGLVLSAAPSPTGVAAT  
QANGPAPAPAPGAPPTGDPKPKKKPKPPRPRAGDNATVAAGHATLREHLR  
DIAKENTDANFVCPPTGATVQFQPRRCPTPRPGASANAVERIKTSSIEPARLOF  
ATMYKYDVTASQVWFGHRSOFGMIFEDAPVPEVVDIKINAKGCRSTAKYVRNL  
ETTAHRDDEHETDMEKAPNAAATRTSRGHTTDLTKNPRVRAFRHGTTCVACIEV  
DASVYPIRNEFLATGDFVYNSMRISRPACYSRPLVSFRYEDQGLIBCOLDENNELR  
LRTDAIEPCTVGRHRRYFTFGGIVYFEERYAHSQSRADITTVSTFIDLTIMLDEH  
PVPLEVYTRHEIKDSGLDYTEVORRNQJHDFADIDIVIHADANAAMFAGLAFEGMG  
DGRVAVKVMGVGVSAVGSFMSNPGALVGLVLAGLVAAPFAFRVYIOL  
QNRPMKALYPLTTKEIKTSDPRGVBGEGBEGAPDCAKLABAREMIRYALVLSAM  
ERTREHAKKKSALLSKVTDMWRKRNTNYTVOPNKGADDEDL"

## BASE COUNT

513 a 951 c 854 g 397 t

Query Match 100.0%; Score 18; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACAGAGCCCGCTGTCTTG 18  
|||||  
Db 2099 ACAGAGCCCGCTGTCTTG 2082

RESULT 15  
AF259899/c 2809 bp DNA linear VRL 16-MAY-2000  
LOCUS

DEFINITION Human herpesvirus 1 glycoprotein B (UL27) gene, complete cds.  
ACCESSION AF259899  
VERSION AF259899.1 GI:7839508  
KEYWORDS  
SOURCE  
ORGANISM Human herpesvirus 1  
Human herpesvirus 1  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE  
AUTHORS 1 (bases 1 to 2809)  
TITLE Ling, J.-Y., Chen, T.-M. and Stroop, W.G.  
A viral genetic element involved in a rabbit model of herpes  
simplex virus-induced epileptiform seizures  
JOURNAL Unpublished  
REFERENCES 2 (bases 1 to 2809)  
AUTHORS Ling, J.-Y., Chen, T.-M. and Stroop, W.G.  
TITLE Direct Submission  
JOURNAL Submitted (24-Apr-2000) Ophthalmology, Mail Slot 523, University of  
Arkansas for Medical Sciences, 4301 West Markham, Little Rock, AR  
72205, USA

FEATURES  
source Location/Qualifiers  
1..2809  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="+GC"  
/db\_xref="taxon:10298"  
38..2752  
/gene="UL27"  
38..2752  
/gene="UL27"  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AAF70301.1"  
/db\_xref="GI:7839509"  
translation="MRQCAPARGCMFVVALIGLTIGLVASAPSSPGTGVAAAT  
QAANGPATPAPAPAPATGPTGTRKAKKPPPPPCGAGNATVAAGHATLREHL  
DIKANTDANFTVCPPTGATVQFEQPRCPTRPGQVTEGLAVFERENTAPYKFK  
ATMYKADVTSQVWFGRYSQPMGI.FEDRAPVPFEVVDKINAGVCRSTACTYRNNL  
ETTAHARDHETDMLKPPANAATRTSRGWHITDLKYNPSRVEAFHRTVNCIVEV  
DARSVPYNEFVLTATGDFYMSAPFYGVRGSHTEHTSYAADRFKOVDFYARDLITKA  
RATAPTRNNILTTPKFTVAMDVPPRPSVCTMTKQVDEMRSEYGSFRSSDAIS  
TTFTTNLTREYPSRYDLSDCIGKARDAMDRI.FARRYNATHIKGQPOYYLANGFPLI  
AVQPLSNSTLAEIYVREHLREQSRKPPNTPPPGASANASVERIKTTSIEFARLQF  
TYNNHQHVNDMLGRVAIAMCLOHNEHLTLMNEARKNPNAIASYVGRVRSARMIGD  
VMAVSTCVPVADNIVIVQNSMRISRPACYSRPLVSFRYEDQPLVEGQLENNEELR  
LTRDAIEPCTVGRHRYFTFGGGVYFEZYVSHOLSRADITTVSTFIDNITMLDHE  
FVPLEVYTRHETKDSGLDYTEVQRNOLHDLRFADIDTVITHADANAAAFAGLGAFFE  
GMGDLGRAVGVKVMGIIVGVSAVSGVSFMSNPFGLAVGLVLAGLAAPFAFRIV  
KRLQSNPMKALVPLTTKELKNTPTNDASGEGEGDFFDAKLAEAREMIRYALVSAM  
BTEHKAKKKGTSLISAKVTDMVRKRNTNYTQVPNKSDADEDDL"

BASE COUNT 536 a 974 c 883 g 416 t  
ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 2809;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 AGCAGCGCGTGTGCTTG 18  
|||  
Db 2136 AGCAGCGCGTGTGCTTG 2119

Search completed: September 16, 2003, 19:14:13  
Job time : 395.752 secs

BEST AVAILABLE COPY

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:02:34 ; Search time 90.1782 Seconds  
(without alignments)  
538.821 Million cell updates/sec

Title: US-09-594-065-2  
Perfect score: 18  
Sequence: 1 agcagggcgcgtgcctg 18

Scoring table: IDENTITY NUC  
Gapop 10-0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18	100.0	2088	12 AAQ14479	Truncated HSVB ge
C 2	18	100.0	2645	19 AAV62158	HSV-2 strain SBS C
C 3	18	100.0	2712	8 AAN71303	Herpes Simplex Vir
C 4	18	100.0	2724	6 AAN71399	Herpes Simplex Vir
C 5	18	100.0	2727	6 AAN50517	Sequence encoding
C 6	18	100.0	2841	19 AAV62146	HSV-2 strain SBS C
C 7	18	100.0	2925	12 AAQ14455	HSV surface antigen
C 8	18	100.0	2943	17 AAT16480	SAB virus GB Glyco

C 9	18	100.0	3098	7 AAN60195	Herpes simplex vir
C 10	18	100.0	3461	8 AAN70681	Recombinant herpes
C 11	18	100.0	3465	12 AAQ14478	HSVB gene. Herpe
C 12	18	100.0	3471	18 AAT93650	Herpes simplex vir
C 13	18	100.0	3472	9 AAN80907	Sequence of Herpes
C 14	18	100.0	3472	14 AAQ48436	Glycoprotein B (gB
C 15	18	100.0	3472	14 AAQ48457	Glycoprotein B (gB
C 16	18	100.0	3473	18 AAT93651	Herpes simplex vir
C 17	18	100.0	3474	9 AAN80908	Sequence of Herpes
C 18	18	100.0	3642	6 AAN50364	DNA fragment encod
C 19	18	100.0	3996	6 AAN50516	Sequence encoding
C 20	18	100.0	117213	19 AAV62176	HSV-2 strain SBS C
C 21	18	100.0	154746	24 AAD25519	Human herpesvirus
C 22	17	94.4	1431	23 AAS54175	Pseudomonas aerugi
C 23	17	94.4	6244	16 AAT06872	PhCMV*-1 promoter
C 24	17	94.4	6244	16 AAO76269	PhCMV*-1-controlle
C 25	17	94.4	6244	20 AAX41724	CDNA encoding rabb
C 26	17	94.4	6244	20 AAX7905	Rabbit progesteron
C 27	17	94.4	6244	20 AAX01369	Rabbit progesteron
C 28	17	94.4	6244	22 AAD09837	Rabbit progesteron
C 29	17	94.4	6244	22 AAL49543	Rabbit progesteron
C 30	17	94.4	6244	25 ABX15341	Rabbit progesteron
C 31	16.4	91.1	512	19 AAV44840	Clone related to F
C 32	16.4	91.1	761	22 AAK73882	Human immune/haema
C 33	16.4	91.1	1110	20 AAX15192	Nucleotide sequenc
C 34	16.4	91.1	1110	20 AAV81738	Sphingomonas sp. s
C 35	16.4	91.1	1110	22 AAS03154	Sphingomonas desul
C 36	16.4	91.1	1436	23 ABU12221	Drosophila melanog
C 37	16.4	91.1	1502	22 AAD17727	Human macrophage-e
C 38	16.4	91.1	1502	23 AAS74722	DNA encoding novel
C 39	16.4	91.1	1896	25 AAF29475	Human monocyte inh
C 40	16.4	91.1	1896	25 ABX17925	Human monocyte inh
C 41	16.4	91.1	1991	19 AAV44826	PCR-III coding seq
C 42	16.4	91.1	2061	24 ABU65650	Lung cancer relate
C 43	16.4	91.1	2170	22 AAF29472	Human monocyte inh
C 44	16.4	91.1	2170	25 ABX17922	DNA encoding human
C 45	16.4	91.1	2173	25 ABX56306	Human NOV27a CG938

#### ALIGNMENTS

AAQ14479/c	AAQ14479 standard; DNA; 2088 BP.
ID	AAQ14479 standard; DNA; 2088 BP.
AC	AAQ14479;
XX	
XX	25-MAR-2003 (updated)
DT	23-JAN-1992 (first entry)
XX	
DE	Truncated HSVB gene.
XX	
KW	Vaccine; antigen; ss.
OS	Herpes simplex virus.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..2088
FT	/*tag= a
PN	JP03218397-A.
XX	
PD	25-SEP-1991.
XX	
XX	21-JUN-1990;
XX	90JP-0161448.
PR	30-NOV-1989;
PR	89JP-0308941.
PR	22-JUN-1989;
PR	89JP-0158238.
PR	21-JUN-1990;
XX	90JP-0161448.
XX	
PA	(TAKE ) TAKEDA CHEM IND LTD.
XX	

DR WPI; 1991-328397/45.  
 XX P-PSDB; AAR14666.  
 DR HSVGB polypeptide(s) obcd. by recombinant DNA techniques -  
 PT useful as vaccines against HSV and in diagnosis, can be produced  
 PT cheaply and safely.  
 XX  
 PS Claim 3; Fig 7; 24pp; Japanese.  
 CC The sequence encodes a truncated form of the HSVGB polypeptide.  
 CC The recombinant protein can be used to prepare vaccines for  
 CC prophylaxis of HSV infection and for use in diagnostic kits.  
 CC See also AAQ14478.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 SQ Sequence 2088 BP; 417 A; 751 C; 620 G; 300 T; 0 other;

Query Match 100.0%; Score 18; DB 12; Length 2088;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGCAGCGCGCTGCTCTTG 18  
 |||||  
 DB 2012 AGCAGCGCGCTGCTCTTG 1995

RESULT 2  
 ID AAV62158 standard; DNA; 2645 BP.  
 AC AAV62158;  
 DT 18-DEC-1998 (first entry)  
 XX  
 DE HSV-2 strain SBS Contig ID 14 DNA sequence.  
 XX  
 KM HSV-2 strain SBS; immunological response induction; therapy;  
 KM antiviral identification; viral protein inhibitor; ss.  
 XX  
 OS Herpes simplex virus type 2.  
 XX

Key Location/Qualifiers  
 CDS complement (81..2645)  
 FT /tag= a  
 FT /product= "ORF1 protein"  
 XX  
 PN MO9820016-A1.  
 XX  
 PD 14-MAY-1998.  
 XX  
 PF 31-OCT-1997; 97WO-US20016.  
 XX  
 PR 09-JUN-1997; 97US-0048018.  
 PR 04-NOV-1996; 96US-0030279.  
 XX  
 PA (SMIK) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;  
 PI Esser KM, Leary JY;  
 XX  
 XX WPI; 1998-286847/25.  
 DR P-PSDB; AAW72113.  
 XX  
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
 PT and treatment of infection or inducing immunological response in  
 PT mammal  
 XX  
 PS Claim 1; Page 406-407; 748pp; English.  
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA  
 CC sequence of the invention. This sequence was isolated from HSV-2 strain  
 CC SBS (deposited as ATCC VR-2546), is designated Contig ID 14, and encodes  
 CC a HSV-2 protein. The protein can be used for the treatment or

CC prevention of disease, to induce an immunological response in a mammal or  
 CC to identify inhibitors, activators or novel antivirals. Antagonists of  
 CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence  
 CC or a vector containing it can also be used to induce an immunological  
 CC response in a mammal.  
 XX  
 SQ Sequence 2645 BP; 404 A; 837 C; 892 G; 512 T; 0 other;

Query Match 100.0%; Score 18; DB 19; Length 2645;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGCAGCGCGCTGCTCTTG 18  
 |||||  
 DB 706 AGCAGCGCGCTGCTCTTG 723

RESULT 3  
 ID AAN71303/C  
 ID AAN71303 standard; DNA; 2712 BP.  
 AC AAN71303;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 30-APR-1991 (first entry)  
 XX  
 DE Herpes Simplex Virus-1 GB from pKBXX.  
 XX  
 KM Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;  
 KM glycoprotein; gb; ss.  
 XX  
 OS Herpes simplex virus type 1 (HOS).  
 XX  
 PN US4642333-A.  
 XX  
 PD 10-FEB-1987.  
 XX  
 PF 20-JUN-1984; 84US-0622496.  
 XX  
 PR 20-JUN-1984; 84US-0622496.  
 XX  
 PA (PERS) PERSON S.  
 XX  
 PI Person S;  
 XX  
 DR WPI; 1987-056354/08.  
 DR P-PSDB; AAP71135.  
 XX  
 PT Amino acid chain of glycoprotein B of HSV-1 and 2 - prepd. as  
 PT recombinant and used for vaccines for herpes simplex virus types 1  
 PT and 2.  
 XX  
 PS Example; Table 2; 16pp; English.  
 XX  
 CC A pure non-glycosylated amino acid (AA) chain comprising a sequence  
 CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2  
 CC virus which is antigenic to HSV-1 of HSV-2, which contains no more  
 CC than 750 AA residues, and which includes AA residues 135-649  
 CC inclusive is claimed. It can be used to produce vaccines for  
 CC prophylaxis and treatment of HSV-1 and HSV-2.  
 CC See AAN71359 for the HSV-2 sequence.  
 CC  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC  
 SQ Sequence 2712 BP; 519 A; 945 C; 849 G; 397 T; 2 other;

Query Match 100.0%; Score 18; DB 8; Length 2712;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGCAGCGCGCTGCTCTTG 18  
 |||||  
 DB 2096 AGCAGCGCGCTGCTCTTG 2079



RESULT 4  
 AAN71399/c  
 ID AAN71399 standard; DNA; 2724 BP.  
 XX  
 AC AAN71399;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 30-APR-1991 (first entry)  
 XX  
 DE Herpes Simplex Virus-2 gB from p52BXX.  
 XX  
 KM Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;  
 KM glycoprotein; gB; ss.  
 XX  
 OS Herpes simplex virus type 2 HGS2.  
 XX  
 PN US642333-A.  
 XX  
 PD 10-FEB-1987.  
 XX  
 PF 20-JUN-1984; 84US-0622496.  
 XX  
 PR 20-JUN-1984; 84US-0622496.  
 XX  
 PA (PERS/) PERSON S.  
 XX  
 PI Person S;  
 XX  
 DR WPI; 1987-056354/08.  
 DR P-PSDB; AAP71136.  
 XX  
 PT Amino acid chain of glyco:protein B of HSV-1 and 2 - prepd. as  
 PT recombinant and used for vaccines for herpes simplex virus types 1  
 PT and 2.  
 XX  
 PS Example; Table 2; 16pp; English.  
 XX  
 CC 'N' indicated nucleotide to be determined (corresp. to residues  
 CC 17-41 of HSV-1 gB).  
 CC A pure non-glycosylated amino acid (AA) chain comprising a sequence  
 CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2  
 CC virus which is antigenic to HSV-1 of HSV-2, which contains no more  
 CC than 750 AA residues, and which includes AA residues 135-649  
 CC inclusive is claimed. It can be used to produce vaccines for  
 CC prophylaxis and treatment of HSV-1 and HSV-2.  
 CC See AAN71303 for the HSV-1 sequence.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC  
 XX  
 SQ Sequence 2724 BP; 494 A; 911 C; 854 G; 390 T; 75 other;  
 Query Match 100.0%; Score 18; DB 8; Length 2724;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGCAGCGCGCTGCTCTTG 18  
 Db 2099 AGCAGCGCGCTGCTCTTG 2082  
 RESULT 5  
 AAN50517/c  
 ID AAN50517 standard; DNA; 2727 BP.  
 XX  
 AC AAN50517;  
 XX  
 DT 21-NOV-1991 (first entry)  
 XX  
 DE Sequence encoding the glycoprotein B (gB) of Herpes simplex virus  
 DE type 2 (HSV-2).  
 XX  
 KM Vaccine; glycoprotein B; ss.  
 XX

OS Herpes simplex virus type 2.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..2727  
 FT /\*tag= a  
 XX  
 FN EP133063-A.  
 XX  
 PD 13-FEB-1985.  
 XX  
 PP 22-JUN-1984; 84EP-0401312.  
 XX  
 PR 16-SEP-1983; 83US-0532996.  
 PR 23-JUN-1983; 83US-0506986.  
 PR 20-JUN-1984; 84US-0622496.  
 XX  
 PA (PERS/) PERSON S.  
 XX  
 PI Person S;  
 XX  
 DR WPI; 1985-039636/07.  
 XX  
 PT New amino acid chains of glyco-protein B of herpes simplex virus  
 PT - prepd. by recombination DNA methods for use in vaccines  
 XX  
 PS Disclosure; Table 2, Page 35-35C; 40pp; English.  
 XX  
 CC The inventors claim a non-glycosylated amino acid chain comprising a  
 CC sequence corresponding to that occurring in glycoprotein B of HSV-1  
 CC or HSV-2. Preferably, when the chain comprises a sequence corresp.  
 CC to that occurring in glycoprotein B of HSV-2, it contains 4-750 AA  
 CC residues. The chain esp. has a molecular wt. of 65000 daltons and  
 CC includes AA residues 165-629 from the HSV-2 glycoprotein B sequence.  
 XX  
 SQ Sequence 2727 BP; 493 A; 913 C; 855 G; 391 T; 75 other;  
 Query Match 100.0%; Score 18; DB 6; Length 2727;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGCAGCGCGCTGCTCTTG 18  
 Db 2102 AGCAGCGCGCTGCTCTTG 2085  
 RESULT 6  
 AAV62146/c  
 ID AAV62146 standard; DNA; 2841 BP.  
 XX  
 AC AAV62146;  
 XX  
 DT 07-DEC-1998 (first entry)  
 XX  
 DE HSV-2 strain SBS Contig ID 91 DNA sequence.  
 XX  
 KM HSV-2 strain SBS; immunological response induction; therapy;  
 KM antiviral identification; viral protein inhibitor; ss.  
 XX  
 OS Herpes simplex virus type 2.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 358..2745  
 FT /\*tag= a  
 XX  
 FN M09820016-A1.  
 XX  
 PD 14-MAY-1998.  
 XX  
 PP 31-OCT-1997; 97WO-US20016.  
 PR 09-JUN-1997; 97US-0049018.  
 PR 04-NOV-1996; 96US-0030279.  
 XX

PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;  
 PI Esser KM, Leary JJ;  
 XX WPI; 1998-286847/25.  
 DR P-PSDB; AAM72062.  
 XX  
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
 PT and treatment of infection or inducing immunological response in  
 PT mammal  
 XX  
 PS Claim 1; Page 284-285; 748pp; English.  
 XX  
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA  
 CC sequence of the invention. This sequence was isolated from HSV-2 strain  
 CC SB5 (deposited as ATCC VR-2546), is designated Contig ID 91, and encodes  
 CC a HSV-2 protein. The protein can be used for the treatment or  
 CC prevention of disease, to induce an immunological response in a mammal or  
 CC to identify inhibitors, activators or novel antivirals. Antagonists of  
 CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence  
 CC or a vector containing it can also be used to induce an immunological  
 CC response in a mammal.  
 CC  
 SQ Sequence 2841 BP; 577 A; 921 C; 915 G; 425 T; 3 other;  
 XX  
 Query Match 100.0%; Score 18; DB 19; Length 2841;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 AGCAGCGCGCTGTCTTG 18  
 DB 2120 AGCAGCGCGCTGTCTTG 2103  
 XX  
 RESULT 7  
 AA014455/C  
 ID AA014455 standard; DNA; 2925 BP.  
 XX  
 AC AA014455;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 21-JAN-1992 (first entry)  
 XX  
 DE HSV surface antigens gB/gD fusion protein gene.  
 XX  
 KM Herpes simplex virus; vaccine; ss.  
 XX  
 OS Herpes simplex virus.  
 XX  
 PH Key Location/Qualifiers  
 FT CDS 1..2925  
 FT /\*tag= a  
 FT  
 FT  
 XX  
 PN JP0320200-A.  
 XX  
 PD 27-SEP-1991.  
 XX  
 PF 29-NOV-1990; 90JP-0325474.  
 XX  
 PR 30-NOV-1989; 89JP-0308942.  
 PR 29-NOV-1990; 90JP-0325474.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 XX  
 DR WPI; 1991-329235/45.  
 DR P-PSDB; AAR14680.  
 XX  
 PT Polypeptide for herpes simplex virus vaccine prodn. - obd. by  
 PT culturing transformant of recombinant DNA coding fused protein of  
 PT virus surface antigen  
 XX  
 PS Claim 6; Fig 8; 16pp; Japanese.

XX  
 CC The DNA sequence encodes a fusion protein comprising herpes simplex  
 CC virus (HSV) surface antigens gB and gD. This fusion protein shows  
 CC the same physiological activity as surface antigen prep. from HSV  
 CC infected cells. It can be used in a vaccine for the prevention of  
 CC HSV infection.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 2925 BP; 579 A; 1061 C; 848 G; 437 T; 0 other;  
 XX  
 Query Match 100.0%; Score 18; DB 12; Length 2925;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 AGCAGCGCGCTGTCTTG 18  
 DB 2012 AGCAGCGCGCTGTCTTG 1995  
 XX  
 RESULT 8  
 AAT16480/C  
 ID AAT16480 standard; DNA; 2943 BP.  
 XX  
 AC AAT16480;  
 XX  
 DT 11-MAY-1996 (first entry)  
 XX  
 DE SAb virus gB glycoprotein coding sequence.  
 XX  
 KM Herpes simian monkey SAb virus gB glycoprotein; immunoassay;  
 KM diagnosis; herpes B virus; ss.  
 XX  
 OS Herpes simian monkey SAb virus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 87..2744  
 FT /\*tag= a  
 FT /product= gB\_glycoprotein  
 XX  
 PN US5487969-A.  
 XX  
 PD 30-JAN-1996.  
 XX  
 PF 01-APR-1993; 93US-0042747.  
 XX  
 PR 01-APR-1993; 93US-0042747.  
 XX  
 PA (SWBI-) SOUTHWEST FOUND BIOMEDICAL RES.  
 XX  
 PI Black D, Eberle R, Hilliard J, Scintariello F;  
 XX  
 DR WPI; 1996-105220/11.  
 DR P-PSDB; AAR92747.  
 XX  
 PT Detection of herpes B virus by PCR amplification of sample DNA - to  
 PT detect a specific herpes simian monkey B virus DNA segment.  
 XX  
 PS Disclosure; Column 23-30; 22pp; English.  
 XX  
 CC This sequence encodes the herpes simian monkey SAb virus gB  
 CC glycoprotein. Such sequences have been used to develop synthetic  
 CC DNA primers which have homologous sequences of conserved regions  
 CC which flank a divergent region of the gB glycoprotein gene. In  
 CC addition, proteins such as the gB glycoprotein have potential use in  
 CC the development of serological immunoassays. One approach is to  
 CC synthesize peptides which, based on the properties of the predicted  
 CC protein sequence, are likely to be immunologically active. Such  
 CC peptides can be used as substrate antigens in immunoassays to detect  
 CC serum antibodies which recognize this specific peptide sequence.  
 CC Synthetic peptides may also be used to produce antibodies against  
 CC specific regions of the gB glycoprotein which are unique to one  
 CC virus. These can then be used to develop virus-specific  
 CC immunoassays for differentiation of SAb virus from other primate

CC alpha-herpes viruses and for identification of antibodies directed  
 CC against SAb virus in primate serum samples.  
 XX  
 SQ Sequence 2943 BP; 462 A; 1103 C; 983 G; 395 T; 0 other;

Query Match 100.0%; Score 18; DB 17; Length 2943;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCAGGCGGCTGTCTTG 18  
 |||||  
 DB 2140 AGCAGGCGGCTGTCTTG 2123

RESULT 9  
 AAN60195/c  
 ID AAN60195 standard; DNA; 3098 BP.

AC AAN60195;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 31-JUL-1991 (first entry)

DE Herpes simplex virus glycoprotein gB gene.

XX HSV; gB glycoprotein; vaccine; ss.

XX Herpes simplex virus.

XX Key Location/Qualifiers  
 FT CDS 257..296  
 /\*tag= a

PN EPI170169-A.

PD 05-FEB-1986.

PF 19-JUL-1985; 85EP-0109042.

PR 20-JUL-1984; 84JP-0151766.  
 11-DEC-1984; 84JP-0262465.

XX (KAGA ) CHERO SERO THERAPEUTIC RES INST.

PI Nozaki C, Makizumi K, Kino Y, Eto, T, Ohtomo N;

DR WPI; 1986-036935/06.  
 P-PSDB; AAP60244.

XX Recombinant DNA containing herpes simplex virus gene or fragment -  
 PT useful in transformant yeast for prodn. of high purity herpes  
 PT simplex virus glycoproteins for use in protective vaccines.

XX Disclosure; Fig. 6; 53pp; English.

XX The gene encodes a gB glycoprotein which is useful for the  
 CC production of vaccines conferring protection against herpes simplex  
 CC virus infections. The vaccine is safe in use and does not contain a  
 CC carcinogenic gene, as with prior vaccines.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 3098 BP; 586 A; 1070 C; 959 G; 483 T; 0 other;

Query Match 100.0%; Score 18; DB 7; Length 3098;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCAGGCGGCTGTCTTG 18  
 |||||  
 DB 2352 AGCAGGCGGCTGTCTTG 2335

RESULT 10  
 AAN70681/c  
 ID AAN70681 standard; DNA; 3461 BP.

AC AAN70681;

XX 25-MAR-2003 (updated)  
 DT 10-MAR-2003 (updated)  
 DT 30-APR-1991 (first entry)

DE Recombinant herpes simplex virus gene.

XX Herpes simplex virus; E.coli; plasmid; vector; promoter; antigen;  
 XX vaccine; glycoprotein; HSV gB; ss.

XX Herpes simplex virus.

XX Key Location/Qualifiers  
 FT CDS 254..2962  
 /\*tag= a  
 /\*product= HSV\_gB

PN JPE2115288-A.

PD 26-MAY-1987.

PF 14-NOV-1985; 85JP-0255971.

PR 14-NOV-1985; 85JP-0255971.

XX (KAGA ) KAGAKU OYOBI KESSEI RYOHU.

DR WPI; 1987-183377/26.  
 P-PSDB; AAP70426.

XX Recombinant DNA contg. herpes simplex virus gene - is integrated in  
 FT E.coli plasmid vector, contg. incorporated promoter region for  
 FT animal cells, giving antigen.

XX Disclosure; Fig 3(1-3); 13pp; Japanese.

XX The gene encoding HSV gB is cloned into E.coli. To increase the  
 CC efficiency of expression, the gene encoding the C-terminal region  
 CC of gB, which is very hydrophobic, is removed and the remaining gene  
 CC is cloned into an expression vector. The vector is composed of pBR322,  
 CC SV40 early promoter and the HSV gB gene.

CC The construct is useful for the production of a good antigen for  
 CC a HSV vaccine.  
 CC (Updated on 10-MAR-2003 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 3461 BP; 666 A; 1188 C; 1041 G; 566 T; 0 other;

Query Match 100.0%; Score 18; DB 8; Length 3461;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCAGGCGGCTGTCTTG 18  
 |||||  
 DB 2349 AGCAGGCGGCTGTCTTG 2332

RESULT 11

AAQ14478/c  
 ID AAQ14478 standard; DNA; 3465 BP.

AC AAQ14478;

XX 25-MAR-2003 (updated)  
 DT 23-JAN-1992 (first entry)

DE HSVgB gene.

```

KW Vaccine; antigen; ss.
XX
XX Herpes simplex virus.
OS
XX
XX Key Location/Qualifiers
XX CDS 251..295
XX /*tag= a
XX
XX JP03218397-A.
XX
XX
XX 25-SEP-1991.
XX
XX 21-JUN-1990; 90JP-0161448.
XX
XX 30-NOV-1989; 89JP-0308941.
XX 22-JUN-1989; 89JP-0158238.
XX 21-JUN-1990; 90JP-0161448.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX WPI; 1991-328397/45.
XX P-P8DB; AAR14665.
XX
XX HSVGB polypeptide(s) obt'd. by recombinant DNA techniques -
XX useful as vaccines against HSV and in diagnosis, can be produced
XX cheaply and safely.
XX
XX PS Claim 3; Fig 1; 24pp; Japanese.
XX
XX CC The sequence encodes the HSVGB polypeptide. The recombinant protein
XX or esp. truncated forms of it, can be used to prepare vaccines for
XX prophylaxis of HSV infection and for use in diagnostic kits.
XX See also AA014479.
XX CC (updated on 25-MAR-2003 to correct PA field.)
XX
XX SQ Sequence 3465 BP, 666 A; 1194 C; 1044 G; 561 T; 0 other;
XX
XX Query Match 100.0%; Score 18; DB 12; Length 3465;
XX Best Local Similarity 100.0%; Pred. No. 34;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AGCAGGCCGCTGCTCTTG 18
XX |||||
XX DB 2349 AGCAGGCCGCTGCTCTTG 2332
XX
XX RESULT 12
XX AAT93650/c
XX ID AAT93650 standard; DNA; 3471 BP.
XX
XX AC AAT93650;
XX
XX DT 26-FEB-1998 (first entry)
XX
XX DE Herpes simplex virus type 1 glycoprotein B encoding DNA.
XX
XX KW Herpes simplex virus; HSV; type I; type II; glycoprotein B; vaccine;
XX membrane protein; virus-specific glycoprotein;
XX transmembrane anchor region; ss.
XX
XX OS Herpes simplex virus type 1.
XX
XX XX Key Location/Qualifiers
XX PH CAAT_signal 55..62
XX FT /*tag= a
XX
XX FT TATA_signal 125..131
XX FT /*tag= b
XX FT /*note= "putative"
XX CDS 309..3023
XX FT /*tag= c
XX FT polyA_signal 3166..3173
XX FT /*tag= d
XX

```

```

FT FT /note= "putative"
FT FT polyA_signal 3409..3416
FT FT /*tag= e
FT FT /*note= "putative"
XX
XX US5648079-A.
XX
XX 15-JUL-1997.
XX
XX 08-DEC-1994; 94US-0351875.
XX
XX 20-OCT-1986; 86US-0921730.
XX 06-APR-1984; 84US-0597784.
XX 20-SEP-1990; 90US-0587179.
XX 21-DEC-1992; 92US-0993415.
XX 18-OCT-1993; 93US-0138717.
XX
XX (CHIR ) CHIRON CORP.
XX
XX PA Burke RL, Pachl C, Valenzuela PDT;
XX
XX PI WPI; 1997-372022/34.
XX
XX DR P-P8DB; AAW34552.
XX
XX DR Vaccines against herpes simplex virus infection - containing
XX recombinant HSV glycoprotein B
XX
XX PS Disclosure; Fig 4A-F; 33pp; English.
XX
XX CC The present sequence is from the Herpes simplex virus (HSV) type I
XX strain Patton and encodes a glycoprotein B. HSV is a double stranded
XX virus packaged within an icosahedral nucleocapsid enveloped within a
XX membrane. The membrane includes a number of virus-specific
XX glycoproteins, with glycoprotein B being one of the most abundant.
XX Glycoprotein B from both HSV type I and type II are cross reactive. New
XX vaccines comprising recombinantly produced glycosylated glycoprotein B
XX that has a deletion of at least part of the transmembrane anchor region,
XX in combination with a carrier and an adjuvant have been produced. The
XX vaccines are used for immunising humans against HSV (HSV type 1 or
XX HSV type 2) by vaccination before or after primary infection with HSV.
XX
XX SQ Sequence 3471 BP, 639 A; 1201 C; 1069 G; 562 T; 0 other;
XX
XX Query Match 100.0%; Score 18; DB 18; Length 3471;
XX Best Local Similarity 100.0%; Pred. No. 34;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 AGCAGGCCGCTGCTCTTG 18
XX |||||
XX DB 2398 AGCAGGCCGCTGCTCTTG 2381
XX
XX RESULT 13
XX AAN80907/c
XX ID AAN80907 standard; DNA; 3472 BP.
XX
XX AC AAN80907;
XX
XX DT 25-MAR-2003 (updated)
XX 09-MAR-1992 (first entry)
XX
XX DE Sequence of Herpes simplex virus (HSV) glycoprotein B (gB) 1 DNA.
XX
XX KW Vaccine; herpes simplex virus; therapy; ss.
XX
XX OS Herpes simplex virus.
XX
XX XX Key Location/Qualifiers
XX PH CDS 309..3023
XX FT /*tag= a
XX
XX PN WO8802634-A.
XX

```

PD 21-APR-1988.  
 XX  
 XX 20-OCT-1987; 87WO-US02709.  
 PF  
 XX 20-JUL-1987; 87US-0079605.  
 PR  
 PR 20-OCT-1986; 86US-0921213.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 PI Burke LR, Pachl C, Valenzuela P;  
 XX  
 DR WPI; 1988-119368/17.  
 DR P-PSDB; AAP80914.  
 XX  
 XX Vaccine for treatment of herpes simplex virus - contains  
 PT recombinant HSV glyco:proteins B and D  
 XX  
 PS Example; Fig 4; 71pp; English.  
 XX  
 CC Prepn. of recombinant GB and GD is described in WO88504587. The  
 CC amino acid sequence and DNA sequence for GB 1 presented in AAP80914  
 CC and AAN80907 differ from that originally presented in Table 1 of  
 CC International Publication No. WO 85/04587. The DNA sequence in Table  
 CC 1 contains an error in that an additional nucleotide (G) is listed  
 CC at position 607 which resulted in a shift in reading frame relative  
 CC to AAN80907 from which this nucleotide has been deleted.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 3472 BP; 639 A; 1202 C; 1068 G; 563 T; 0 other;  
 Query Match 100.0%; Score 18; DB 9; Length 3472;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGCAGCGCGCTGCTCTTG 18  
 DB 2398 AGCAGCGCGCTGCTCTTG 2381

RESULT 14  
 AAQ48496/C  
 ID AAQ48496 standard; DNA; 3472 BP.  
 XX  
 AC AAQ48496;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 29-MAR-1994 (first entry)  
 XX  
 DE Glycoprotein B (GB1).  
 XX  
 KW Glycoprotein; GB1; GB2; herpes simplex virus; HSV-1; HSV-2;  
 KW flanking; initiation; termination; transcription; translation;  
 KW regulatory sequence; ss.  
 XX  
 OS Herpes simplex virus 1 strain Patton.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 309..3023  
 FT /\*tag= a  
 FT  
 XX  
 PN US5244792-A.  
 XX  
 PD 14-SEP-1993.  
 PD  
 PF 20-SEP-1990; 90US-0587179.  
 PF  
 XX 06-APR-1984; 84US-0597784.  
 PR 20-OCT-1986; 86US-0921730.  
 PR 20-SEP-1990; 90US-0587179.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Burke RL, Pachl C, Valenzuela PDT;

XX  
 DR WPI; 1993-302641/38.  
 DR P-PSDB; AAR41778.  
 XX  
 XX DNA construct for expressing HSV glycosylated polypeptide -  
 PT useful for vaccinating against HSV-1 and -2 infections  
 XX  
 XX Disclosure; Fig 4; 33pp; English.  
 XX  
 CC DNA constructs for expressing a glycosylated polypeptide in a  
 CC eukaryotic host cell comprises (1) an oligonucleotide sequence (OS)  
 CC free of natural flanking sequences, encoding glycoprotein B (GB)  
 CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating  
 CC transcriptional and translational regulatory sequences flanking OS,  
 CC at least one of these sequences not being from HSV. The HSV-1 and  
 CC HSV-2 sequences are given in AAQ48496 and AAQ48497 respectively.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 3472 BP; 639 A; 1202 C; 1069 G; 562 T; 0 other;  
 Query Match 100.0%; Score 18; DB 14; Length 3472;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGCAGCGCGCTGCTCTTG 18  
 DB 2398 AGCAGCGCGCTGCTCTTG 2381

RESULT 15  
 AAQ48497/C  
 ID AAQ48497 standard; DNA; 3472 BP.  
 XX  
 AC AAQ48497;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 29-MAR-1994 (first entry)  
 XX  
 DE Glycoprotein B (GB2).  
 XX  
 KW Glycoprotein; GB1; GB2; herpes simplex virus; HSV-1; HSV-2;  
 KW flanking; initiation; termination; transcription; translation;  
 KW regulatory sequence; ss.  
 XX  
 OS Herpes simplex virus 2 strain 333.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 309..3023  
 FT /\*tag= a  
 FT  
 XX  
 PN US5244792-A.  
 XX  
 PD 14-SEP-1993.  
 PD  
 PF 20-SEP-1990; 90US-0587179.  
 PF  
 XX 06-APR-1984; 84US-0597784.  
 PR 20-OCT-1986; 86US-0921730.  
 PR 20-SEP-1990; 90US-0587179.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Burke RL, Pachl C, Valenzuela PDT;  
 XX  
 DR WPI; 1993-302641/38.  
 DR P-PSDB; AAR41779.  
 XX  
 XX DNA construct for expressing HSV glycosylated polypeptide -  
 PT useful for vaccinating against HSV-1 and -2 infections  
 XX  
 XX Disclosure; Fig 4; 33pp; English.  
 XX  
 CC DNA constructs for expressing a glycosylated polypeptide in a

CC eukaryotic host cell sequences (1) an oligonucleotide sequence (OS)  
CC fire of natural flanking sequences, encoding glycoprotein B (GB)  
CC of HSV, or its fragments and (2) 5' initiating and 3' terminating  
CC transcriptional and translational regulatory sequences flanking OS  
CC at least one of these sequences not being from HSV. The HSV-1 and  
CC HSV-2 sequences are given in AAQ4896 and AAQ4847 respectively.  
CC (Updated on 25-MAR-2005 to correct Pf field.)

80 Sequence 3472 BP; 661 A; 1193 C; 1053 G; 565 T; 0 other;

Query Match	Score 18;	DB 14;	Length 3472;
Best local similarity	100.0%		
Best local similarity	100.0%	Prod No. 34.	

Matches 18; Conservative 0; Mismatches

18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGGCCGCTGTCCTTG 18  
|||  
Db 2407 AGCAGGCCGCTGTCCTTG 2390

2407 AGCAGGCCGCTGTCCTTG 2390

Search completed: September 16, 2003, 18:36:58  
Job time : 91.1782 secs

Job time : 91.1782 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 18:24:35 ; Search time 22.9901 Seconds  
(without alignments)  
345.579 Million cell updates/sec

Title: US-09-594-065-2

Perfect score: 18

Sequence: 1 agcagcgccgtccttg 18

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues 1139956

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	2713	US-08-804-439A-13	Sequence 13, Appl
2	18	100.0	2713	US-08-720-229-13	Sequence 13, Appl
3	18	100.0	2943	US-08-042-747A-7	Sequence 7, Appl
4	18	100.0	3472	5244792-1	Patent No. 5244792
5	18	100.0	3472	5244792-2	Patent No. 5244792
6	17	94.4	855	US-09-252-991A-1201	Sequence 1201, Ap
7	17	94.4	1491	US-09-252-991A-1243	Sequence 1243, Ap
8	17	94.4	1551	US-09-252-991A-1362	Sequence 1362, Ap
9	17	94.4	6244	US-08-076-726-15	Sequence 15, Appl
10	17	94.4	6244	US-08-260-452-8	Sequence 8, Appl
11	17	94.4	6244	US-08-481-970-8	Sequence 8, Appl
12	17	94.4	6244	US-08-897-718-8	Sequence 8, Appl
13	17	94.4	6244	US-09-163-269-8	Sequence 8, Appl
14	17	94.4	711	US-09-252-991A-12195	Sequence 12195, A
15	16.4	91.1	1110	US-08-851-088-9	Sequence 9, Appl
16	16.4	91.1	1149	US-09-252-991A-12362	Sequence 12362, A
17	16.4	91.1	1251	US-09-252-991A-12303	Sequence 12303, A
18	16.4	91.1	1251	US-09-252-991A-12121	Sequence 12121, A
19	16.4	91.1	1251	US-09-345-468-24	Sequence 24, Appl
20	16.4	91.1	1896	US-09-414-453A-24	Sequence 24, Appl
21	16.4	91.1	1896	US-09-832-312-24	Sequence 24, Appl
22	16.4	91.1	2170	US-09-345-468-11	Sequence 11, Appl
23	16.4	91.1	2170	US-09-414-453A-11	Sequence 11, Appl
24	16.4	91.1	2170	US-09-832-312-11	Sequence 11, Appl
25	16.4	91.1	2170	US-09-310-463-19	Sequence 19, Appl
26	16.4	91.1	2194	US-08-842-248A-19	Sequence 19, Appl
27	16.4	91.1	2194		

C 28	16.4	91.1	3177	1	US-08-042-747A-4	Sequence 4, Appl
C 29	16.4	91.1	4144	3	US-08-851-089-12	Sequence 12, Appl
C 30	16.4	91.1	4144	3	US-08-851-089-13	Sequence 13, Appl
C 31	16	88.9	2244	1	US-08-476-519-10	Sequence 10, Appl
C 32	16	88.9	2244	5	PCT-US95-09323-10	Sequence 10, Appl
C 33	16	88.9	2244	1	US-08-476-519-1	Sequence 1, Appl
C 34	16	88.9	2334	5	PCT-US95-09323-1	Sequence 1, Appl
C 35	15.4	85.6	3014	1	US-08-629-939-1	Sequence 1, Appl
C 36	15.4	85.6	3014	1	US-08-759-873-1	Sequence 1, Appl
C 37	15.4	85.6	3014	4	US-09-016-434-1244	Sequence 1244, Ap
C 38	15.4	85.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 39	15.4	85.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 40	14.8	82.2	290	4	US-09-313-294A-6122	Sequence 6122, Ap
C 41	14.8	82.2	471	4	US-09-252-991A-263	Sequence 263, App
C 42	14.8	82.2	516	4	US-09-252-991A-15658	Sequence 15658, A
C 43	14.8	82.2	684	4	US-09-252-991A-11110	Sequence 11110, A
C 44	14.8	82.2	747	3	US-08-998-416-246	Sequence 246, App
C 45	14.8	82.2	765	4	US-09-199-637A-324	Sequence 324, App

## ALIGNMENTS

RESULT 1  
US-08-804-439A-13/c  
Sequence 13, Application US/08804439A  
Patent No. 6015565  
GENERAL INFORMATION:  
APPLICANT: Rose, Timothy M.  
APPLICANT: Bosch, Martin L.  
TITLE OF INVENTION: GLYCOPROTEIN B OF THE RHY/KSHV  
TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Fish & Richardson P.C.  
STREET: 4225 Executive Square, Ste 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804, 439A  
FILING DATE: February 21, 1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Hallie, Lisa A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 09176/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070  
TELEFAX: (619) 678-5099  
TELEX:  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2713 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-804-439A-13  
Query Match 100.0%; Score 18; DB 3; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGCAGCGCGCTGCTCTTG 18  
DB 2096 AGCAGCGCGCTGCTCTTG 2079

```
RESULT 2
US-08-720-229-13/C
; Sequence 13, Application US/08720229
; Patent No. 6022542
; GENERAL INFORMATION:
; APPLICANT: Rose, Timothy M.
; APPLICANT: Boesch, Marnix L.
; APPLICANT: Strand, Kurt
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RRV/KSHV
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES
; NUMBER OF SEQUENCES: 100
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/720,229
; FILING DATE: 26-SEP-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Schiffe, J. Michael
; REGISTRATION NUMBER: 40,253
; REFERENCE/DOCKET NUMBER: 29938-20002.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-720-229-13
Query Match      100.0%; Score 18; DB 3; Length 2713;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCAGGCCGCTGTCTTG 18
      |||||
Db      2096 AGCAGGCCGCTGTCTTG 2079

RESULT 3
US-08-042-747A-7/C
; Sequence 7, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scintiscariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; TITLE OF INVENTION: Virus Genes
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
```

```
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042,747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, M. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..2744
;
US-08-042-747A-7
Query Match      100.0%; Score 18; DB 1; Length 2943;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCAGGCCGCTGTCTTG 18
      |||||
Db      2140 AGCAGGCCGCTGTCTTG 2123

RESULT 4
5244792-1/C
; Patent No. 5244792
; APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.
; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYOPROTEIN
; B FROM HERPES SIMPLEX VIRUS
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,179
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 921,730
; FILING DATE: 20-OCT-1986
; APPLICATION NUMBER: 597,784
; FILING DATE: 06-APR-1984
; SEQ ID NO:1:
; LENGTH: 3472
;
5244792-1
Query Match      100.0%; Score 18; DB 6; Length 3472;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCAGGCCGCTGTCTTG 18
      |||||
Db      2398 AGCAGGCCGCTGTCTTG 2381

RESULT 5
5244792-2/C
; Patent No. 5244792
; APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.
; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYOPROTEIN
; B FROM HERPES SIMPLEX VIRUS
; NUMBER OF SEQUENCES: 19
```



```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/587,179
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 921,730
; FILING DATE: 20-OCT-1986
; APPLICATION NUMBER: 597,784
; FILING DATE: 06-APR-1984
; SEQ ID NO: 2:
; LENGTH: 3472
5244792-2

Query Match          100.0%; Score 18; DB 6; Length 3472;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCAGCGCGCTGTCTCTTG 18
Db      2407 AGCAGCGCGCTGTCTCTTG 2390

RESULT 6
US-09-252-991A-1201/c
; Sequence 1201, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1201
; LENGTH: 855
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1201

Query Match          94.4%; Score 17; DB 4; Length 855;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GCAGCGCGCTGTCTCTTG 18
Db      539 GCAGCGCGCTGTCTCTTG 523

RESULT 7
US-09-252-991A-1243/c
; Sequence 1243, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1243
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1243

Query Match          94.4%; Score 17; DB 4; Length 1491;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GCAGCGCGCTGTCTCTTG 18
Db      424 GCAGCGCGCTGTCTCTTG 408

RESULT 8
US-09-252-991A-1362
; Sequence 1362, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1362
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1362

Query Match          94.4%; Score 17; DB 4; Length 1551;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GCAGCGCGCTGTCTCTTG 18
Db      1116 GCAGCGCGCTGTCTCTTG 1132

RESULT 9
US-08-076-726-15/c
; Sequence 15, Application US/08076726
; Patent No. 5464758
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Tight Control of Gene Expression in
; TITLE OF INVENTION: Eucaryotic Cells by Tetracycline-responsive Promoters
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/076,726
; FILING DATE: 14-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0942, 2490001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
```

TELEFAX: (202) 371-2540  
TELEX: 248636 SSK  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6244 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
US-08-076-726-15

Query Match 94.4%; Score 17; DB 1; Length 6244;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 17  
|||||  
DB 991 AGCAGCGCGCTGTCTT 975

## RESULT 10

US-08-260-452-8/c  
Sequence 8, Application US/08260452  
Patent No. 5650298  
GENERAL INFORMATION:  
APPLICANT: Goosen, Manfred  
APPLICANT: Bujard, Hermann  
APPLICANT: Salfeld, Jochen  
APPLICANT: Voss, Jeffrey  
TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic  
TITLE OF INVENTION: Cells by Tetracycline-responsive Promoters  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Lahive & Cockfield  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/260,452  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/076,327  
ATTORNEY/AGENT INFORMATION:  
NAME: Giulio A. DeConti, Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-013CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6244 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Human cytomagalovirus  
STRAIN: Towne (hCMV)  
IMMEDIATE SOURCE:  
CLONE: pUHD BGR3  
US-08-260-452-8

Query Match 94.4%; Score 17; DB 1; Length 6244;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 17  
|||||  
DB 991 AGCAGCGCGCTGTCTT 975

## RESULT 11

US-08-481-970-8/c  
Sequence 8, Application US/08481970  
Patent No. 5859310  
GENERAL INFORMATION:  
APPLICANT: Goosen, Manfred  
APPLICANT: Bujard, Hermann  
APPLICANT: Salfeld, Jochen  
APPLICANT: Voss, Jeffrey  
TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Lahive & Cockfield  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,970  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/260,452  
FILING DATE: 14-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/076,327  
FILING DATE: 14-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Giulio A. DeConti, Jr.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: BBI-013CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6244 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Human cytomagalovirus  
STRAIN: Towne (hCMV)  
IMMEDIATE SOURCE:  
CLONE: pUHD BGR3  
US-08-481-970-8

Query Match 94.4%; Score 17; DB 2; Length 6244;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 17  
|||||  
DB 991 AGCAGCGCGCTGTCTT 975

## RESULT 12

US-08-897-719-8/c  
Sequence 8, Application US/08897719  
Patent No. 5922927

Transcript:

```

; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Tight Control of Gene Expression in Eucaryotic
; TITLE OF INVENTION: Cells by Tetracycline-Responsive Promoters
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahnive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,719
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/260,452
; FILING DATE:
; APPLICATION NUMBER: 08/076,327
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; STRAIN: Towne (hcmv)
; IMMEDIATE SOURCE:
; CLONE: PUHD BGR3
; US-08-897-719-8

Query Match          94.4%; Score 17; DB 2; Length 6244;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCAGCGCGCTGTCTT 17
Db      991 AGCAGCGCGCTGTCTT 975

RESULT 13
US-09-163-269-8/c
; Sequence 8, Application US/09163269
; Patent No. 6252136
; GENERAL INFORMATION:
; APPLICANT: Gossen, Manfred
; APPLICANT: Bujard, Hermann
; APPLICANT: Salfeld, Jochen
; APPLICANT: Voss, Jeffrey
; TITLE OF INVENTION: Animal Transgenic for a Tetracycline-Controlled
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahnive & Cockfield
; STREET: 60 State Street, Suite 510
; Transcripts
```

```

; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,269
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/481,970
; FILING DATE:
; APPLICATION NUMBER: 08/076,327
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/076,327
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. DeConti, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: BBI-013CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Human cytomegalovirus
; STRAIN: Towne (hcmv)
; IMMEDIATE SOURCE:
; CLONE: PUHD BGR3
; US-09-163-269-8

Query Match          94.4%; Score 17; DB 3; Length 6244;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCAGCGCGCTGTCTT 17
Db      991 AGCAGCGCGCTGTCTT 975

RESULT 14
US-09-252-991A-12195/c
; Sequence 12195, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12195
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-12195

Query Match          91.1%; Score 16.4; DB 4; Length 711;
Best Local Similarity 94.4%; Pred. No. 36;
```

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTTG 18  
 |||||  
 DB 514 AGCAGCGCGTGTCTTG 497

## RESULT 15

US-08-851-088-9  
 / Sequence 9, Application US/08851088  
 / Patent No. 5952208  
 / GENERAL INFORMATION:  
 / APPLICANT: Darzine, Aldis  
 / APPLICANT: X1, Lei  
 / APPLICANT: Childs, John D.  
 / APPLICANT: Monticello, Daniel J.  
 / APPLICANT: Squires, Charles H.  
 / TITLE OF INVENTION: DSZ Gene Expression In Pseudomonas Hosts  
 / NUMBER OF SEQUENCES: 23  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.  
 / STREET: Two Militia Drive  
 / CITY: Lexington  
 / STATE: Massachusetts  
 / COUNTRY: USA  
 / ZIP: 02173  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent In Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/851,088  
 / FILING DATE: 05-May-1997  
 / CLASSIFICATION: 435  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/835,185  
 / FILING DATE: 07-APR-1997  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Elmore, Carolyn S.  
 / REGISTRATION NUMBER: 37,567  
 / REFERENCE/DOCKET NUMBER: EBC96-06A  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (781) 861-6240  
 / TELEFAX: (781) 861-9540  
 / INFORMATION FOR SEQ. ID NO: 9:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1110 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: DNA (genomic)  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: 1..1107  
 / US-08-851-088-9

Query Match 91.1%; Score 16.4; DB 2; Length 1110;

Best Local Similarity 94.4%; Pred. No. 36;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGTGTCTTG 18  
 |||||  
 DB 620 AGCAGCGCGTGTCTTG 637

Search completed: September 16, 2003, 20:42:36  
 Job time : 27.9901 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:14:25 ; Search time 65.0495 Seconds

(without alignments)  
680.690 Million cell updates/sec

Title: US-09-594-065-2

Perfect score: 18

Sequence: 1 agcagggccgctgccttg 18

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1660708 seqs, 122959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_NEW\_PUB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	154746	12	US-09-827-688-8
2	17	94.4	1431	9	US-09-815-242-7812
3	17	94.4	6244	9	US-09-281-674-8
4	17	94.4	6244	10	US-09-892-227-8
5	16.4	91.1	609	13	US-10-027-632-12963
6	16.4	91.1	609	13	US-10-027-632-12964
7	16.4	91.1	1896	9	US-09-832-312-24
8	16.4	91.1	2061	10	US-09-954-456-960
9	16.4	91.1	2170	9	US-09-832-312-11
10	16.4	91.1	2194	12	US-10-143-618-19
11	16.4	91.1	2194	14	US-10-139-662-19
12	16.4	91.1	2194	14	US-10-139-683-19
13	16.4	91.1	13553	11	US-09-764-891-7432
14	16	88.9	976	14	US-10-198-846-11382
15	15.4	85.6	489	10	US-09-764-868-222
16	15.4	85.6	573	13	US-10-027-632-15146

C 17	15.4	85.6	622	13	US-10-027-632-267576	Sequence 267576,
C 18	15.4	85.6	2190	14	US-10-156-761-2735	Sequence 2735, Ap
C 19	15.4	85.6	2453	10	US-09-974-298-111	Sequence 111, App
C 20	15.4	85.6	2470	9	US-09-870-937-9	Sequence 9, Appl
C 21	15.4	85.6	2470	10	US-09-954-456-1121	Sequence 1121, Ap
C 22	15.4	85.6	2470	10	US-09-880-107-1725	Sequence 1725, Ap
C 23	15.4	85.6	2470	12	US-10-354-358-7	Sequence 7, Appl
C 24	15.4	85.6	2519	13	US-10-071-766-82	Sequence 82, Appl
C 25	15.4	85.6	2799	9	US-09-935-368-2	Sequence 2, Appl
C 26	15.4	85.6	5003	14	US-10-207-655-62	Sequence 62, Appl
C 27	15.4	85.6	23098	14	US-10-017-161-2005	Sequence 2005, Ap
C 28	15.4	85.6	9025608	14	US-10-156-761-1	Sequence 1, Appl
C 29	15	83.3	262	9	US-09-250-883-1	Sequence 1, Appl
C 30	15	83.3	263	9	US-09-250-883-2	Sequence 2, Appl
C 31	15	83.3	1332	9	US-09-250-883-14	Sequence 14, Appl
C 32	15	83.3	1929	14	US-10-106-698-805	Sequence 805, App
C 33	15	83.3	3446	12	US-09-769-734-4	Sequence 4, Appl
C 34	14.8	82.2	60	12	US-09-908-975-13289	Sequence 13289, A
C 35	14.8	82.2	262	9	US-09-923-876-2088	Sequence 2088, Ap
C 36	14.8	82.2	371	10	US-09-796-692-6249	Sequence 6249, Ap
C 37	14.8	82.2	371	14	US-10-040-662-6249	Sequence 6249, Ap
C 38	14.8	82.2	424	13	US-10-001-857-58	Sequence 58, Appl
C 39	14.8	82.2	439	10	US-09-796-692-2556	Sequence 2556, Ap
C 40	14.8	82.2	439	14	US-10-040-862-2556	Sequence 2556, Ap
C 41	14.8	82.2	471	11	US-09-918-995-32269	Sequence 32269, A
C 42	14.8	82.2	478	13	US-10-027-632-287388	Sequence 287388,
C 43	14.8	82.2	618	12	US-10-308-779-67	Sequence 67, Appl
C 44	14.8	82.2	765	11	US-09-975-719-324	Sequence 324, App
C 45	14.8	82.2	810	14	US-10-156-761-2123	Sequence 2123, Ap

## ALIGNMENTS

RESULT 1  
US-09-827-688-8  
; Sequence 8, Application US/09827688  
; Publication No. US20030165476A1  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BERMA  
; APPLICANT: BHOQAL, BALBIR  
; TITLE OF INVENTION: MACROGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION D  
; FILE REFERENCE: P01949US1/10004014  
; CURRENT APPLICATION NUMBER: US/09/827,688  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8

Query Match 100.0%; Score 18; DB 12; Length 154746;  
Best Local Similarity 100.0%; Pred. No. 3.4;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGGCGCTGCTTG 18  
DB 54028 AGCAGGCGCTGCTTG 54045

RESULT 2  
US-09-815-242-7812/c  
; Sequence 7812, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.

```

1  APPLICANT: Zykend, Judith W.
2  APPLICANT: Wall, Daniel
3  APPLICANT: Trawick, John D.
4  APPLICANT: Carr, Grant J.
5  APPLICANT: Yamamoto, Robert T.
6  APPLICANT: Xu, H. Howard
7  TITLE OF INVENTION: Identification of Essential Genes in
8  TITLE OF INVENTION: Prokaryotes
9  FILE REFERENCE: ELITRA.011A
10 CURRENT APPLICATION NUMBER: US/09/815,242
11 CURRENT FILING DATE: 2001-03-21
12 PRIOR APPLICATION NUMBER: 60/191,078
13 PRIOR FILING DATE: 2000-03-21
14 PRIOR APPLICATION NUMBER: 60/206,848
15 PRIOR FILING DATE: 2000-05-23
16 PRIOR APPLICATION NUMBER: 60/207,727
17 PRIOR FILING DATE: 2000-05-26
18 PRIOR APPLICATION NUMBER: 60/242,578
19 PRIOR FILING DATE: 2000-10-23
20 PRIOR APPLICATION NUMBER: 60/253,625
21 PRIOR FILING DATE: 2000-11-27
22 PRIOR APPLICATION NUMBER: 60/257,931
23 PRIOR FILING DATE: 2000-12-22
24 PRIOR APPLICATION NUMBER: 60/269,308
25 PRIOR FILING DATE: 2001-02-16
26 NUMBER OF SEQ ID NOS: 1410
27 SOFTWARE: FastSeq for Windows Version 4.0
28 SEQ ID NO 7812
29
30 LENGTH: 1431
31 TYPE: DNA
32 ORGANISM: Pseudomonas aeruginosa
33 FEATURE:
34 NAME/KEY: CDS
35 LOCATION: (1)...(1431)
36 US-09-815-242-7812

```

```

Query Match Similarity 94.4%; Score 17; DB 9; Length 1431;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Cy 2 GCAGGCCGCTGTCCTTG 18
|||||
|||||
Db 364 GCAGGCCGCTGTCCTTG 348

RESULT 3
US-09-281-674-8/c
: Sequence 8, Application US/09281674
: Patent No. US20020077307A1
: GENERAL INFORMATION:
: APPLICANT: Goosen, Manfred
: Bujard, Hermann
: Salfeld, Jochen
: Voos, Jeffrey
: TITLE OF INVENTION: Methods for Regulating Gene Expression
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lahive & Cockfield
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/281,674
: FILING DATE: 30-Mar-1999
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
:

```

```

1  APPLICATION NUMBER: 08/479,306
2  FILING DATE: <Unknown>
3  APPLICATION NUMBER: 08/260,452
4  FILING DATE: 14-JUN-1994
5  APPLICATION NUMBER: 08/076,327
6  FILING DATE: 14-JUN-1993
7  ATTORNEY/AGENT INFORMATION:
8  NAME: Giulio A. Decont'i, Jr.
9  REGISTRATION NUMBER: 31,503
10 REFERENCE/DOCKET NUMBER: BBI-013CPB
11
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (617) 227-7430
14 TELEFAX: (617) 227-5941
15
16 INFORMATION FOR SEQ ID NO: 8:
17
18 SEQUENCE CHARACTERISTICS:
19
20     LENGTH: 6244 base pairs
21     TYPE: nucleic acid
22     STRANDEDNESS: double
23     TOPOLOGY: circular
24
25     MOLECULE TYPE: DNA (genomic)
26
27     ORIGINAL SOURCE:
28
29         ORGANISM: Human cytomegalovirus
30         STRAIN: Towne (hcmv)
31
32     IMMEDIATE SOURCE:
33         CLONE: pUD6 BGR3
34
35     SEQUENCE DESCRIPTION: SEQ ID NO: 8:
36
37 JS-09-281-674-8

```

	94.4%; Score 17; DB 9; Length 6244;
	Best Local Similarity 100.0%; Pred. No. 17;
Matches	17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 AGCAGCGCCGCTGTCCTT 17
db	991 AGCAGCGCCGCTGTCCTT 975

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

REFERENCE/DOCKET NUMBER: BBI-013CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6244 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Human cytomegalovirus  
STRAIN: Towne (hcmv)  
IMMEDIATE SOURCE:  
CLONE: PUHD BGR3  
SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-892-227-8

Query Match 94.4%; Score 17; DB 10; Length 6244;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 17  
|||||  
Db 991 AGCAGCGCGCTGTCTT 975

RESULT 5  
US-10-027-632-12963  
; Sequence 12963, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108627.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 12963  
; LENGTH: 609  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-12963

Query Match 91.1%; Score 16.4; DB 13; Length 609;  
Best Local Similarity 94.4%; Pred. No. 45;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 18  
|||||  
Db 97 AGCAGCGCGCTGTCTT 114

RESULT 6  
US-10-027-632-12964  
; Sequence 12964, Application US/10027632  
; GENERAL INFORMATION:

APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108627.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 12964  
LENGTH: 609  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-12964

Query Match 91.1%; Score 16.4; DB 13; Length 609;  
Best Local Similarity 94.4%; Pred. No. 45;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 18  
|||||  
Db 97 AGCAGCGCGCTGTCTT 114

RESULT 7  
US-09-832-312-24/c  
; Sequence 24, Application US/09832312  
; Patent No. US20010049829A1  
; GENERAL INFORMATION:  
; APPLICANT: Busfield et al.  
; TITLE OF INVENTION: GLYCOPROTEIN VI AND USES THEREOF  
; FILE REFERENCE: 7853-234  
; CURRENT APPLICATION NUMBER: US/09/832,312  
; CURRENT FILING DATE: 2001-04-09  
; PRIOR APPLICATION NUMBER: 09/610,118  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 09/503,387  
; PRIOR FILING DATE: 2000-02-14  
; PRIOR APPLICATION NUMBER: 09/454,824  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 09/345,468  
; PRIOR FILING DATE: 1999-06-30  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO: 24  
; LENGTH: 1896  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-832-312-24

Query Match 91.1%; Score 16.4; DB 9; Length 1896;  
Best Local Similarity 94.4%; Pred. No. 39;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTT 18  
|||||  
Db 1496 AGCAGCGCGCTGTCTT 1479

RESULT 8

```

US-09-954-456-960/c
Sequence 960, Application US/09954456
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
PRIORITY FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: Patentin version 3.0
SEQ ID NO 960
LENGTH: 2061
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-960

Query Match          91.1%   Score 16.4; DB 10; Length 2061;
Best Local Similarity 94.4%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGCAGGCCGCTGTCTTGG 18
        |||||
Db       1496 AGCAGGCCCTGTCTTG 1479

RESULT 9
US-09-832-312-11/c
Sequence 11, Application US/09832312
Patent No. US20010049829A1
GENERAL INFORMATION:
APPLICANT: Buelfield et al.
TITLE OF INVENTION: GLYCOPROTEIN VI AND USES THEREOF
FILE REFERENCE: 7853-234
CURRENT APPLICATION NUMBER: US/09/832,312
CURRENT FILING DATE: 2001-04-09
PRIOR APPLICATION NUMBER: 09/610,118
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 09/503,387
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 09/454,824
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 09/345,468
PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 78
SOFTWARE: fastseq for windows Version 3.0
SEQ ID NO 11
LENGTH: 2170
TYPE: DNA
ORGANISM: Homo sapiens
US-09-832-312-11
```

```

Query Match          91.1%; Score 16.4; DB 9; Length 2170;
Best Local Similarity 94.4%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGCAGCCGCTGTCCTTG 18
      |||||
Db       1545 AGCAGCCCTGTCCTTG 1528

RESULT 10
US-10-143-618-19/c
; Sequence 19, Application US/10143618
; Publication No. US20030166073A1
; GENERAL INFORMATION:
APPLICANT: Cosman, David J.
TITLE OF INVENTION: Family of Immunoregulators Designated
Leukocyte Immunoglobulin-Like Receptors (LIR)
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSER: Janis C. Henry, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM/PC Compatible
OPERATING SYSTEM: Microsoft Word 7.0
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/143,618
FILING DATE: 08-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/942,248A
FILING DATE: April 24, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Henry, Janis C.
REGISTRATION NUMBER: 34,347
REFERENCE/DOCKET NUMBER: 2624
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2194 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: phm17
FEATURE:
NAME/KEY: CDS
LOCATION: 67..1959
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-10-143-618-19

Query Match          91.1%; Score 16.4; DB 12; Length 2194;
Best Local Similarity 94.4%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGCAGCCGCTGTCCTTG 18
      |||||
Db       1562 AGCAGCCCTGTCCTTG 1545

RESULT 11
US-10-139-662-19/c
; Sequence 19, Application US/10139662
; Publication No. US20030027358A1

```



```

; GENERAL INFORMATION:
; APPLICANT: Cosman, David J.
; APPLICANT: Anderson, Dirk M.
; APPLICANT: Borges, Luis
; TITLE OF INVENTION: Family of Immunoregulators Designated Leukocyte Immunoglobulin-
; TITLE OF INVENTION: Like Receptors (LIR)
; FILE REFERENCE: 2624-A
; CURRENT APPLICATION NUMBER: US/10/139,662
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: 09/310,463
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 08/842,248
; PRIOR FILING DATE: 1997-04-24
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2194
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67)..(1962)
US-10-139-662-19

Query Match          91.1%; Score 16.4; DB 14; Length 2194;
Best Local Similarity 94.4%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCCTTG 18
Db 1562 AGCAGCGCGCTGTCCTTG 1545

RESULT 12
US-10-139-683-19/C
; Sequence 19, Application US/10139683
; Publication No. US20030060614A1
; GENERAL INFORMATION:
; APPLICANT: Cosman, David J.
; APPLICANT: Anderson, Dirk M.
; APPLICANT: Borges, Luis
; TITLE OF INVENTION: Family of Immunoregulators Designated Leukocyte Immunoglobulin-
; TITLE OF INVENTION: Like Receptors (LIR)
; FILE REFERENCE: 2624-A
; CURRENT APPLICATION NUMBER: US/10/139,683
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: 09/310,463
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/842,248
; PRIOR FILING DATE: 1997-04-24
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2194
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (67)..(1962)
US-10-139-683-19

Query Match          91.1%; Score 16.4; DB 14; Length 2194;
Best Local Similarity 94.4%; Pred. No. 38;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCCTTG 18
Db 1562 AGCAGCGCGCTGTCCTTG 1545
```

```

; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7432
; LENGTH: 13553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7432

Query Match          91.1%; Score 16.4; DB 11; Length 13553;
Best Local Similarity 94.4%; Pred. No. 30;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCCTTG 18
Db 1984 AGCAGCGCTGTCCTTG 1967

RESULT 14
US-10-198-846-11382/C
; Sequence 11382, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11382
; LENGTH: 976
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-846-11382

Query Match          89.9%; Score 16; DB 14; Length 976;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGGCCGCTGTCCTTG 18
Db 589 CAGGCCGCTGTCCTTG 574

RESULT 15
US-09-764-868-222
; Sequence 222, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 222
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 18:22:25 ; Search time 911.049 Seconds  
(without alignments)  
480.194 Million cell updates/sec

Title: US-09-594-065-2

Perfect score: 1 agcagcgcgcgcgcctcg 18

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estcp1:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_dln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.4	91.1	195	9	AM663294 bh77a10.y
2	16.4	91.1	287	14	T63918 yc08c12.f1
3	16.4	91.1	308	28	A2081267 SLCA2A2.Fe
4	16.4	91.1	356	13	BY310214 BY310214

5	16.4	91.1	364	13	BY664919
6	16.4	91.1 <td>366</td> <td>13 <td>BY064920</td> </td>	366	13 <td>BY064920</td>	BY064920
7	16.4	91.1 <td>373</td> <td>13 <td>BY661558</td> </td>	373	13 <td>BY661558</td>	BY661558
8	16.4	91.1 <td>388</td> <td>13 <td>BY066019</td> </td>	388	13 <td>BY066019</td>	BY066019
9	16.4	91.1 <td>391</td> <td>9</td> <td>AV661443</td>	391	9	AV661443
10	16.4	91.1 <td>424</td> <td>13 <td>BY223280</td> </td>	424	13 <td>BY223280</td>	BY223280
11	16.4	91.1 <td>442</td> <td>13 <td>BQ317016</td> </td>	442	13 <td>BQ317016</td>	BQ317016
12	16.4	91.1 <td>522</td> <td>10</td> <td>BE490435</td>	522	10	BE490435
13	16.4	91.1 <td>546</td> <td>10</td> <td>BG633702</td>	546	10	BG633702
14	16.4	91.1 <td>554</td> <td>28</td> <td>AQ488864</td>	554	28	AQ488864
15	16.4	91.1 <td>598</td> <td>13</td> <td>BQ240732</td>	598	13	BQ240732
16	16.4	91.1 <td>600</td> <td>14</td> <td>CB553914</td>	600	14	CB553914
17	16.4	91.1 <td>623</td> <td>28</td> <td>AZ321116</td>	623	28	AZ321116
18	16.4	91.1 <td>713</td> <td>9</td> <td>A1296995</td>	713	9	A1296995
19	16.4	91.1 <td>822</td> <td>10</td> <td>BE410139</td>	822	10	BE410139
20	16.4	91.1 <td>847</td> <td>13</td> <td>BUI34166</td>	847	13	BUI34166
21	16.4	91.1 <td>875</td> <td>13</td> <td>B0937820</td>	875	13	B0937820
22	16.4	91.1 <td>898</td> <td>13</td> <td>BX453678</td>	898	13	BX453678
23	16.4	91.1 <td>913</td> <td>29</td> <td>B2578934</td>	913	29	B2578934
24	16.4	91.1 <td>940</td> <td>29</td> <td>B2571854</td>	940	29	B2571854
25	16.4	91.1 <td>1198</td> <td>29</td> <td>B2556562</td>	1198	29	B2556562
26	16.4	91.1 <td>2432</td> <td>11</td> <td>BC028153</td>	2432	11	BC028153
27	16.4	91.1 <td>271</td> <td>28</td> <td>BH641766</td>	271	28	BH641766
28	16.4	91.1 <td>345</td> <td>28</td> <td>B2187917</td>	345	28	B2187917
29	16.4	91.1 <td>417</td> <td>14</td> <td>CB767279</td>	417	14	CB767279
30	16.4	91.1 <td>470</td> <td>28</td> <td>BH320515</td>	470	28	BH320515
31	16.4	91.1 <td>481</td> <td>14</td> <td>CB728483</td>	481	14	CB728483
32	16.4	91.1 <td>500</td> <td>29</td> <td>B2285951</td>	500	29	B2285951
33	16.4	91.1 <td>536</td> <td>28</td> <td>BH328864</td>	536	28	BH328864
34	16.4	91.1 <td>630</td> <td>13</td> <td>B0671252</td>	630	13	B0671252
35	16.4	91.1 <td>647</td> <td>28</td> <td>B2177926</td>	647	28	B2177926
36	16.4	91.1 <td>656</td> <td>10</td> <td>BP289102</td>	656	10	BP289102
37	16.4	91.1 <td>674</td> <td>14</td> <td>CA504539</td>	674	14	CA504539
38	16.4	91.1 <td>711</td> <td>10</td> <td>BG347557</td>	711	10	BG347557
39	16.4	91.1 <td>801</td> <td>29</td> <td>B2283539</td>	801	29	B2283539
40	16.4	91.1 <td>808</td> <td>13</td> <td>B0255259</td>	808	13	B0255259
41	16.4	91.1 <td>817</td> <td>14</td> <td>CD458065</td>	817	14	CD458065
42	16.4	91.1 <td>1064</td> <td>10</td> <td>BE540693</td>	1064	10	BE540693
43	16.4	91.1 <td>212</td> <td>10</td> <td>BF804235</td>	212	10	BF804235
44	16.4	91.1 <td>285</td> <td>14</td> <td>CD265227</td>	285	14	CD265227
45	16.4	91.1 <td>300</td> <td>9</td> <td>AU098914</td>	300	9	AU098914

## ALIGNMENTS

RESULT 1  
AM663294/c 195 bp mRNA linear EST 06-APR-2000  
bh77a10.y1 NCI CGAP GUI Homo sapiens CDNA clone IMAGE:2968794.5'  
similar to TR:015471 015471 MONOCYTE INHIBITORY RECEPTOR PRECURSOR.  
// mRNA sequence.

ACCESSION AM663294.1 GI:7455836  
VERSION AM663294  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 195)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished  
COMMENT Other ESTs: bh77a10.x1  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
Technologies, Inc. CDNA Library Arrayed by: Chistea Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:  
 image.lnl.gov/image/html/resources.shtml  
 Seq primer: -40RP from Glibco  
 High quality sequence stop: 193.

## FEATURES

## source

1..195  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2968794"  
 /cissue\_type="2 pooled high-grade transitional cell  
 tumors"  
 /lab\_host="DH10B"  
 /clone\_id="NCI CGAP GU1"  
 /note="Organ: gynecological tract; Vector: pCMV-SPORT6;  
 Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally.  
 Primer: Oligo dT. Library constructed by Life  
 Technologies."

BASE COUNT 54 a 50 c 62 g 25 t 4 others  
 ORIGIN

Query Match 91.1%; Score 16.4; DB 9; Length 195;  
 Best Local Similarity 94.4%; Pred. No. 2.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCAGGCGCTGCTCTTG 18  
 |||||  
 85 ACCAGGCGCTGCTCTTG 68

RESULT 2  
 T63918/c 287 bp mRNA linear EST 17-FEB-1995  
 LOCUS YC08C12.1 Stratiogene lung (#937210) Homo sapiens CDNA clone  
 IMAGE:80086 5' similar to SP:RRT2\_PAT P06768 RETINOL-BINDING  
 PROTEIN II, CELLULAR; mRNA sequence.

ACCESSION T63918  
 VERSION T63918.1 GI:667783  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM

REFERENCE  
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M. P., Chappel, B.,  
 Chiswick, S., Dietrich, N., Dubuque, T., Favallo, A., Gish, W., Hawkins,  
 M., Hutteman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,  
 B., Morris, M., Parsons, J., Prange, C., Riklin, L., Rohlfing, T.,  
 Schellendberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Treva, E.,  
 Underwood, K., Wohldmann, P., Waterston, R., Wilson, R. and Marra, M.  
 Generation and analysis of 280,000 human expressed sequence tags

TITLE  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 MEDLINE 57044478  
 PUBMED 8889549

## COMMENT

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: ead@wustl.wustl.edu  
 Insert Size: 779  
 High quality sequence stops: 271 Source: IMAGE Consortium, LNL. This  
 clone is available royalty-free through LNL; contact the IMAGE  
 Consortium (info@image.lnl.gov) for further information.  
 Insert Length: 779 Std Error: 0.00  
 Seq primer: M13R1  
 High quality sequence stop: 271.  
 Location/Qualifiers

## FEATURES

## source

1..287  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="GDB:483703"  
 /db\_xref="taxon:9606"

/clone="IMAGE:80086"  
 /sex="male"  
 /dev stage="72 years"  
 /lab\_host="SOLR cells (kenamycin resistant)"  
 /clone\_id="Stratiogene lung (#937210)"  
 /note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI  
 ; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
 dT, normal lung. Average insert size: 1.0 kb; Uni-ZAP XR  
 Vector; -5' adaptor sequence: 5' GAAATCGGACGACG 3' -3'  
 adaptor sequence: 5' CTCAGCTTTT TTTT TTTT 3' "

BASE COUNT 65 a 78 c 71 g 53 t  
 ORIGIN

Query Match 91.1%; Score 16.4; DB 14; Length 287;  
 Best Local Similarity 94.4%; Pred. No. 2.5e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCAGGCGCTGCTCTTG 18  
 |||||  
 62 ACCAGGCGCTGCTCTTG 45

RESULT 3  
 AZ081267/c 308 bp DNA linear GSS 23-AUG-2000  
 LOCUS SLC2A2 Felis catus fca215 Felis catus genomic, genomic survey  
 DEFINITION  
 ACCESSION AZ081267  
 VERSION AZ081267.1 GI:7405601  
 KEYWORDS GSS.  
 SOURCE Felis catus (cat)  
 ORGANISM

REFERENCE  
 AUTHORS Murphy, M. D., Sun, S., Chen, Z.-Q., Yuhki, N., Hirschmann, D.,  
 Menotti-Raymond, M., and O'Brien, S. J.  
 A radiation hybrid map of the cat genome: implications for  
 comparative mapping  
 Genome Res. 10 (5), 691-702 (2000)  
 JOURNAL 20272149  
 MEDLINE 10810092  
 PUBMED 10810092

## FEATURES

## source

1..308  
 /organism="Felis catus"  
 /mol\_type="Genomic DNA"  
 /strain="Fca215"  
 /db\_xref="taxon:9685"  
 /sex="male"  
 /clone\_id="Felis catus fca215"

BASE COUNT 57 a 71 c 75 g 105 t  
 ORIGIN

Query Match 91.1%; Score 16.4; DB 28; Length 308;  
 Best Local Similarity 94.4%; Pred. No. 2.6e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCAGGCGCTGCTCTTG 18  
 |||||  
 95 ACCAGGCGCTGCTCTTG 78

RESULT 4  
 BY310214 356 bp mRNA linear EST 11-DEC-2002  
 LOCUS BY310214

DEFINITION BY10214 RIKEN full-length enriched, stroma cell Mus musculus cDNA clone I320013601 5', mRNA sequence.

ACCESSION BY310214

VERSION BY310214.1 GI:26500551

KEYWORDS EST

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Mikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Walstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyszewski, B., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carinici, P., Hayatsu, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Nakamura, M., Sakazume, N., Sato, K., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL NATURE 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Shubiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Atsawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Takashi Ishikawa (Department of Surgery

2 Yokohama City University 3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 Japan ) whose assistance we gratefully acknowledge. Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source location/Qualifiers

1..356

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="CS7BL/6J"

/db\_xref="taxon:10090"

/clone="I320013601"

/cell\_type="stroma cell"

/clone\_lib="RIKEN full-length enriched, stroma cell"

58 a 108 c 115 g 75 t

BASE COUNT

ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 356;

Best Local Similarity 94.4%; Pred. No. 2.7e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGCAGCGCGTGTCTTG 18

127 ACCAGGCCGCTCTTG 144

Db

RESULT 5

BY664919

LOCUS

DEFINITION BY664919 RIKEN full-length enriched, 14.5 days embryo Rp-/+ Rathe's pouches Mus musculus cDNA clone K72029L07 3', mRNA sequence.

ACCESSION BY664919

VERSION BY664919.1 GI:27039049

KEYWORDS EST

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Mikaido, I., Oatso, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Walstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyszewski, B., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carinici, P., Hayatsu, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Nakamura, M., Sakazume, N., Sato, K., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL NATURE 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sec.riken.go.jp  
URL: http://genome.sec.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane  
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y. Direct  
Substation

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN Integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichipillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept.  
Human Genetics University of Michigan Medical School 4301  
MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA )  
whose assistance we gratefully acknowledge.  
Please visit our web site (http://genome.sec.riken.go.jp) for  
further details.

#### FEATURES

##### source

Location/Qualifiers  
1. 364  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="K720029L07"  
/tissue\_type="Rachke's pouches"  
/dev\_stage="14.5 days embryo RP+/-"  
/clone\_lib="RIKEN full-length enriched, 14.5 days embryo  
RP+/- Rachke's pouches"  
BASE COUNT 108 a 79 c 102 t

Query Match 91.1% Score 16.4; DB 13; Length 364;  
Best Local Similarity 94.4%; Pred. No. 2.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGGCGCTGCTTGG 18  
|||||  
Db 102 AGCAGGCGCTGCTTGG 119

RESULT 6  
BY064920 366 bp mRNA linear EST 06-DEC-2002  
LOCUS BY064920 RIKEN full-length enriched, 17 days embryo heart Mus  
DEFINITION musculus cDNA clone I920038D20 5', mRNA sequence.  
ACCESSION BY064920  
VERSION BY064920.1 GI:26169005  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 366)  
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,

#### TITLE

JOURNAL  
MEDLINE  
PUBMED

#### COMMENT

Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S.,  
Bisiel, K. W., Blake, J. A., Brad, D., Brusic, V., Chotila, C., Corbett,  
L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest,  
A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,  
Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedler, R. N.,  
King, B. L., Konegaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,  
P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki,  
H., Nagashima, T., Numata, K., Okido, T., Paven, W. J., Pettes, G.,  
Pescio, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,  
Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,  
B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Semple, C. A., Seton,  
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,  
R. D., Tomita, M., Verardo, R., Wagner, L., Wabstet, C., Wang, Y.,  
Watanabe, Y., Wells, C., Wilming, L. G., Wyshaw-Boris, A., Yanagisawa,  
M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,  
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Haashizume, W., Imotani, K., Ishii,  
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

12466851

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sec.riken.go.jp,  
URL: http://genome.sec.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane  
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y. Direct  
Substation

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN Integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichipillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (http://genome.sec.riken.go.jp) for  
further details.

#### FEATURES

##### source

Location/Qualifiers  
1. 366  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I920038D20"  
/tissue\_type="heart"  
/dev\_stage="17 days embryo"  
/clone\_lib="RIKEN full-length enriched, 17 days embryo  
heart"  
BASE COUNT 84 a 103 c 91 g 87 t 1 others  
ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 366;  
 Best Local Similarity 94.4%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGCAGCGCGCTGCTCTTG 18  
 |||||  
 25 AGCAGCGCGCGCTGCTTG 42

RESULT 7  
 BY661558  
 LOCUS BY661558 RIKEN full-length enriched, 14.5 days embryo RP+/+  
 DEFINITION Ratche's pouches Mus musculus cDNA clone K720011A10 3', mRNA  
 sequence.  
 BY661558  
 ACCESSION BY661558.1 GI:270331604  
 VERSION EST.  
 KEYWORDS Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 373)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,  
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani,  
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Glessi, C., Godzik, A.,  
 Gough, J., Grimond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,  
 Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 B.Z., Ringwald, M., Sadelin, A., Schneider, C., Sempke, C.A., Setou,  
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carninci, P., Hayatsu, N., Hirozane-Kikukawa, T., Konno, H., Makamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Atakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
 K., Shingawa, A., Yasuniishi, A., Yoshino, M., Waterston, R., Lander,  
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT

Contact: Yoshinide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Tsurumi-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-resgsc.riken.go.jp/  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Aikawa, T., Carninci, P., Fukuda, S., Hirozane,  
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Komori, H., Miyazaki, A.,  
 Murata, M., Nakamura, M., Nomura, K., Nomura, R., Ohno, M., Sasaki, K.,  
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
 M., Waki, K., Watanabe, A., Watanabe, M. and Hayashizaki, Y. Direct  
 Submission  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multichannel sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in RIKEN  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Michelle Brinkmeier and Sally Camper (Dept.  
 Human Genetics University of Michigan Medical School 4301  
 MSRB 3 1500 W. Medical Center Dr. Ann Arbor, MI 48109-0638 USA )  
 whose assistance we gratefully acknowledge.  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.

FEATURES  
 source  
 Location/Qualifiers  
 1..373  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="K720011A10"  
 /tissue\_type="Ratche's pouches"  
 /dev\_stage="14.5 days embryo RP+/+"  
 /clone\_id="RIKEN full-length enriched, 14.5 days embryo  
 RP+/+ Ratche's pouches"  
 110 a 79 c 79 g 105 t

BASE COUNT  
 ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 373;  
 Best Local Similarity 94.4%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGCAGCGCGCTGCTCTTG 18  
 |||||  
 110 AGCAGCGCGCTGCTCTTG 127

RESULT 8  
 BY666019  
 LOCUS BY666019 388 bp mRNA linear EST 06-DEC-2002  
 DEFINITION BY666019 RIKEN full-length enriched, 17 days embryo heart Mus  
 musculus cDNA clone I920046H19 5', mRNA sequence.  
 BY666019  
 ACCESSION BY666019.1 GI:26169868  
 VERSION EST.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 388)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,  
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani,  
 L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Glessi, C., Godzik, A.,  
 Gough, J., Grimond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,  
 Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 B.Z., Ringwald, M., Sadelin, A., Schneider, C., Sempke, C.A., Setou,  
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa,

**TITLE**  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

**REFERENCE**  
Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carinci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Maki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ichi, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterson, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

**CONTACT**  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/

**ABSTRACT**  
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1571-1771 (2000)  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
source  
Location/Qualifiers  
1..388  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="1920046H19"  
/tissue\_type="heart"  
/dev\_stage="17 days embryo"  
/clone\_lib="RIKEN full-length enriched, 17 days embryo heart"

**BASE COUNT**  
ORIGIN  
66 a 140 c 91 g 88 t 3 others

**Query Match**  
Best Local Similarity 91.1%; Score 16.4; DB 13; Length 388;  
Pred. No. 2.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**DB**  
1 AGCAGCGCGCTGCTTG 18  
|||||  
25 AGCAGCGCGCTGCTTG 42

**RESULT 9**  
LOCUS  
DEFINITION  
AV661443 G1C Homo sapiens cDNA clone G1CGS803 3', mRNA sequence.  
AV661443 391 bp mRNA linear EST 16-JAN-2002  
AV661443

**VERSION**  
AV661443.1 GI:9682457  
EST.

**KEYWORDS**  
SOURCE  
Homo sapiens (human)

**ORGANISM**  
Homo sapiens

**REFERENCE**  
Bukacynski, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 391)  
XU, X., HUANG, J., QIAN, B., ZHU, Z., YAN, Q., CAI, T., ZHANG, X., XIAO, H., GU, J., LIU, F., ZHONG, M., XU, S., GU, W., HUANG, W., ZHAO, X., SHEN, K., LU, G., FU, G., ZHONG, M., XU, S., GU, W., HUANG, W., ZHAO, X., HU, G., GU, J., CHEN, Z., and HAN, Z.  
Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver  
Proc. Natl. Acad. Sci. U.S.A. 98 (126), 15089-15094 (2001)

**CONTACT**  
Contact: Zeguang Han  
Chinese National Human Genome Center at Shanghai  
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China  
Tel: 86-21-50801919(ex.45)  
Fax: 86-21-50801922  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

**FEATURES**  
source  
Location/Qualifiers  
1..391  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="G1CGS803"  
/tissue\_type="corresponding non cancerous liver tissue"  
/dev\_stage="Adult"  
/lab\_host="SOLR"  
/clone\_lib="GLC"  
/note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2: XhoI"

**BASE COUNT**  
ORIGIN  
89 a 107 c 118 g 77 t

**Query Match**  
Best Local Similarity 91.1%; Score 16.4; DB 9; Length 391;  
Pred. No. 2.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**DB**  
1 AGCAGCGCGCTGCTTG 18  
|||||  
297 AGCAGCGCGCTGCTTG 280

**RESULT 10**  
LOCUS  
DEFINITION  
BY223280 424 bp mRNA linear EST 10-DEC-2002  
BY223280 RIKEN full-length enriched, activated spleen Mus musculus cDNA clone F830302M06 5', mRNA sequence.  
BY223280  
BY223280.1 GI:26404383  
EST.

**KEYWORDS**  
SOURCE  
Mus musculus (house mouse)

**ORGANISM**  
Mus musculus

**REFERENCE**  
Bukacynski, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 424)  
OKAZAKI, Y., FURUNO, M., KASUKAWA, T., ADACHI, J., BONO, H., KONDO, S., MIKAI, I., OZATO, N., SATO, R., SUZUKI, H., YAMANAKA, I., KIYOSAWA, H., YAGI, K., TOMARU, Y., HASEGAWA, Y., NOGAMI, A., SCHONBACH, C., GOJOBORI, T., BALDARELLI, R., HILL, D. P., BULT, C., HUNE, D. A., QUACKENBUSH, J., SCHIRIM, L. M., KANAPIN, A., MALANDA, H., BATELOV, S., BEISEL, K. W., BLAKE, J. A., BRADY, D., BRUSC, V., CHOCHIA, C., CORBANI, L. E., COUSINS, S., DALLA, E., DRAGANI, T. A., FLECHER, C. F., FORRESTER, A., FRAZER, K. S., GAESTERLAND, T., GARIBOLDI, M., GLISCI, C., GODZIK, A., GOUGH, J., GRIMMOND, S., GUARINICH, S., HIROKAWA, N., JACKSON, I. J., JARVIS, E. D., KANAI, A., KAWAJI, H., KAWASAWA, Y., KEDZIERSKI, R. M., KING, B. L., KONAGAYA, A., KUROCHKIN, I. V., LEE, Y., LENHARD, B., LYONS



P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontus, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, J. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempke, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavalan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kikukawa, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analyses of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

12466851

22354683

JOURNAL MEDLINE PUBMED

COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-research@riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

Location/Qualifiers

1. .424

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="NOD"

/db\_xref="taxon:10090"

/clone="F830302M05"

/tissue\_type="activated spleen"

/clone\_lib="RIKEN full-length enriched, activated spleen"

BASE COUNT 80 a 127 c 131 g 86 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 424;  
Best Local Similarity 94.4%; Pred. No. 2.8e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGCTCTTG 18  
|||||  
Db 139 AGCAGCGCGCTGCTCTTG 156  
|||||

RESULT 11

LOCUS B0317016 442 bp mRNA linear EST 17-MAY-2002

DEFINITION IL3-CT0214-150200-076-F11 CT0214 Homo sapiens cDNA, mRNA sequence.

ACCESSION B0317016

VERSION B0317016.1 GI:20922785

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 442)

Diag Neto, B., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. P., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

JOURNAL MEDLINE PUBMED

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL3&r2=IL3-CT0214-150200-076-F11&t3=2000-02-15&t4=1>)  
Seq primer: puc 18 forward.

Location/Qualifiers

1. .442

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/dev\_stage="Adult"

/clone\_lib="CT0214"

/note="Organ: colon; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ONESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 104 a 114 c 120 g 104 t

ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 442;  
Best Local Similarity 94.4%; Pred. No. 2.8e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGCTCTTG 18  
|||||  
Db 89 AGCAGCGCGCTGCTCTTG 106  
|||||

RESULT 12

LOCUS BR490435 522 bp mRNA linear EST 31-JUL-2000

DEFINITION WH030367\_D11.G2125 wheat cold-stressed seedling cDNA library

Triticum aestivum cDNA clone wheat WH030367\_D11.G21, mRNA sequence.

ACCESSION BR490435

VERSION BE490435.1 GI:9610064  
 KEYWORDS EST.  
 SOURCE Triticum aestivum (bread wheat)  
 ORGANISM Triticum aestivum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae  
 1 (bases 1 to 522)  
 REFERENCE 1 (bases 1 to 522)  
 AUTHORS Anderson, O.D., Choe, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hala, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L., and Tong, J.C.  
 TITLE The structure and function of the expressed portion of the wheat genome - Cold-stressed seedling cDNA library  
 JOURNAL Unpublished  
 COMMENT Contact: Olin Anderson  
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818  
 Email: oanderson@w.usda.gov  
 Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20  
 Seq primer: StrataGene SX primer.  
 Location/Qualifiers  
 1..522  
 /organism="Triticum aestivum"  
 /mol\_type="rRNA"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="WHE0367.D11.G21"  
 /cfeature="5'UTR"  
 /cfeature="Coding Sequence"  
 /dev\_stage="Five-day old seedling"  
 /lab\_host="E. coli SOLR"  
 /clone\_1lb="Wheat cold-stressed seedling cDNA library"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
 Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, mycstatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give phagescript phagemids in the 10 Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."  
 BASE COUNT 150 a 103 c 140 g 129 t  
 ORIGIN  
 Query Match 91.1%; Score 16.4; DB 10; Length 522;  
 Best Local Similarity 94.4%; Pred. No. 2.9e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGCAGGCCGCTGCTCTTG 18  
 |||||  
 DB 296 AGCAGGCCGCTGCTCTTG 313  
 |||||  
 RESULT 13  
 BG633702/c 546 bp mRNA linear EST 23-APR-2001  
 DEFINITION AT29407 Splice AT Drosophila melanogaster adult testes polyB7  
 Drosophila melanogaster cDNA clone AT29407 5 similar to CG8214:  
 FBan0008214 located on: 2R 52C2-52C2; 04/09/2001, mRNA sequence.  
 ACCESSION BG633702  
 VERSION BG633702.1 GI:13761239  
 KEYWORDS EST.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

REFERENCE Ephydriidae; Drosophilidae; Drosophila.  
 1 (bases 1 to 546)  
 AUTHORS Stapleton, M., Brokstein, P., Hong, L., Aspayani, A., Baxter, E., Berman, B., Carlson, J., Champe, M., Chavez, C., Chew, M., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Miranda, A., Mista, S., Mungall, C.J., Nunoo, J., Pech, J., Parag, V., Park, S., Phuanavong, S., Wan, K., Yu, C., Lewis, S.E., Celis, S., and Rubin, G.M.  
 TITLE BDGP/HMI AT Drosophila EST Project  
 JOURNAL Unpublished  
 COMMENT Contact: Stapleton, M.  
 BDGP  
 Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_estefruitfly.berkeley.edu](http://www.fruitfly.org/EST_estefruitfly.berkeley.edu)  
 hit genomic AE003809; arm:2R [10637169,10898136]  
 estimated-cyto:52C1-52D9; 04/09/2001  
 Plate: AT.294 row: A column: 7  
 High quality sequence scop: 472.  
 Location/Qualifiers  
 1..546  
 /organism="Drosophila melanogaster"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:7227"  
 /clone="AT29407"  
 /sex="male"  
 /dev\_stage="0-3 day old Ore-R males"  
 /lab\_host="Plates AT.10-AT.120: DHS-alpha. Plates AT.121-AT.319: DHS-alpha Tona"  
 /clone\_1lb="AT Drosophila melanogaster adult testes polyB7"  
 /note="Organ: ADULT testes; Vector: polyB7; Site 1: EcoRI; Site 2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly ligated into polyB7. Plasmid cDNA library."  
 BASE COUNT 156 a 138 c 153 g 99 t  
 ORIGIN  
 Query Match 91.1%; Score 16.4; DB 10; Length 546;  
 Best Local Similarity 94.4%; Pred. No. 2.9e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGCAGGCCGCTGCTCTTG 18  
 |||||  
 DB 336 AGCTGGCCGCTGCTCTTG 319  
 |||||  
 RESULT 14  
 AQ488864/c 554 bp DNA linear GSS 24-APR-1999  
 DEFINITION RPCI-11-266G3.TV RPCI-11 Homo sapiens genomic clone RPCI-11-266G3,  
 genomic survey sequence.  
 ACCESSION AQ488864  
 VERSION AQ488864.1 GI:4674738  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE Zhao, S., Adams, M.D., Niernman, W., Malek, J., de Jong, P., and Venter, J.C.  
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
 JOURNAL Unpublished  
 COMMENT Other GSSes: RPCI-11-266G3.TV  
 Contact: Shaying Zhao, William Niernman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200

Fax: 301 838 0208  
 Email: hbeet@igrr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@delong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Gene Co (<http://www.rgco.com>). BAC end search page: [http://www.rgco.com/cdb/human/bac\\_end\\_search/bac\\_end\\_search.html](http://www.rgco.com/cdb/human/bac_end_search/bac_end_search.html).  
 Seq primer: 17  
 Class: BAC ends.

## FEATURES

source

location/Qualifiers

1..554  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7601906"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-266G3"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPCI-11"  
 /note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; RPCI1 Human Male BAC Library"

## BASE COUNT

163 a 148 c 84 g 153 t

## ORIGIN

Query Match 91.1%; Score 16.4; DB 28; Length 554;  
 Best Local Similarity 94.4%; Pred. No. 3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTCTTG 18  
 ||||| ||||| ||||| |||||  
 Db 416 AGCAGCGCGCTGTCTCTTG 399

## RESULT 15

BQ240732

LOCUS BQ240732 598 bp mRNA linear EST 03-MAY-2002  
 DEFINITION TaB05013B12R TaB05 Triticum aestivum CDNA clone TaB05013B12R, mRNA  
 sequence.

## ACCESSION

BQ240732

## VERSION

BQ240732.1 GI:20436608

## KEYWORDS

EST.

## SOURCE

Triticum aestivum (bread wheat)

## ORGANISM

Triticum aestivum

## REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

## AUTHORS

; Triticaceae; Triticum.

## TITLE

1 (bases 1 to 598)

## JOURNAL

Cloutier, S.

## COMMENT

Wheat functional genomics - Glenlea developing seeds cDNA libraries

Unpublished  
 Contact: Dr. Sylvie Cloutier  
 Cereal Research Centre, Agriculture and Agri-food Canada  
 195 Dufour Rd, Winnipeg, MB, Canada R3T 2M9  
 Tel: (204) 983-2340  
 Fax: (204) 983-4604  
 Email: scloutier@agr.ca  
 was cloned directionally, not all sequences generated with reverse  
 primer were from the 5' end (same with forward primer and 3' end).  
 Average insert size is >2.0 kb  
 Plate: 013 row: B column: 12  
 Seq primer: M13 Reverse.

## FEATURES

source

location/Qualifiers

1..598  
 /organism="Triticum aestivum"  
 /mol\_type="mRNA"  
 /cultivar="Glenlea"  
 /db\_xref="taxon:4565"  
 /clone="TaB05013B12R"  
 /tissue\_type="developing seeds"  
 /dev\_stage="5 days after anthesis"  
 /lab\_host="E. coli DH10B"  
 /clone\_lib="TaB05"

/note="Vector: pSPORT-P (Invitrogen Technologies); Site 1:  
 NotI; Site 2: MluI; mRNA obtained from wheat seeds of  
 cultivar Glenlea 5 days post-anthesis"

## BASE COUNT

170 a 121 c 157 g 150 t

## ORIGIN

Query Match 91.1%; Score 16.4; DB 13; Length 598;  
 Best Local Similarity 94.4%; Pred. No. 3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCAGCGCGCTGTCTCTTG 18  
 ||||| ||||| ||||| |||||  
 Db 327 AGCAGCGCGCTGTCTCTTG 344

Search completed: September 16, 2003, 20:39:53  
 U00 time : 915.049 secs

**BEST AVAILABLE COPY**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:35:29 ; Search time 438.614 Seconds  
(without alignments)  
1865.405 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20  
Sequence: 1 tcaccaccgcgcacaccc 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sfs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sfs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_ay:\*  
39: em\_hcgo\_hum:\*  
40: em\_hcgo\_mus:\*  
41: em\_hcgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	798	6 103657	103657 Sequence 5
2	20	100.0	1478	14 S74390	S74390 gBpath/UL27
3	20	100.0	1488	14 S65444	S65444 UL27-glycop
4	20	100.0	1857	14 H51GBM	K02720 HSV1 (mutan
5	20	100.0	2088	6 E03093	E03093 DNA encodin
6	20	100.0	2586	6 E00358	E00358 DNA encodin
7	20	100.0	2712	6 BD145144	BD145144 Herpes si
8	20	100.0	2712	6 BD165793	BD165793 Vaccine f
9	20	100.0	2712	6 BD165795	BD165795 Vaccine f
10	20	100.0	2715	6 BD145143	BD145143 Herpes si
11	20	100.0	2715	14 AF097023	AF097023 Human her
12	20	100.0	2715	14 AF295528	AF295528 Human her
13	20	100.0	2715	14 AF311740	AF311740 Human her
14	20	100.0	2715	14 HHU49121	HHU49121 Human herp
15	20	100.0	2809	14 AF259899	AF259899 Human her
16	20	100.0	2818	14 AF021340	AF021340 Human her
17	20	100.0	2846	14 HHU12172	HHU12172 Human herp
18	20	100.0	2846	14 HHU12174	HHU12174 Human herp
19	20	100.0	2855	14 HHU12173	HHU12173 Human herp
20	20	100.0	2855	14 HHU12175	HHU12175 Human herp
21	20	100.0	2925	6 E03115	E03115 DNA encodin
22	20	100.0	3098	6 E00883	E00883 Fragment of
23	20	100.0	3324	14 HS2GB	M14923 HSV2 glycop
24	20	100.0	3461	6 E01195	E01195 DNA sequenc
25	20	100.0	3465	6 E03024	E03024 DNA encodin
26	20	100.0	3465	6 E03092	E03092 DNA encodin
27	20	100.0	3465	6 E03112	E03112 DNA encodin
28	20	100.0	3472	6 BD165792	BD165792 Vaccine f
29	20	100.0	3472	6 BD165794	BD165794 Vaccine f
30	20	100.0	3472	6 I08685	I08685 Sequence 1
31	20	100.0	3472	6 I08686	I08686 Sequence 2
32	20	100.0	3472	14 HS2GB3	M5118 HSV-2 (333)
33	20	100.0	3640	6 I08361	I08361 Sequence 1
34	20	100.0	3643	14 HS1GBP	K03541 HSV-1 (Pat
35	20	100.0	3755	6 E03025	E03025 DNA encodin
36	20	100.0	3755	6 E03113	E03113 DNA encodin
37	20	100.0	3758	14 HS1GB	K01760 HSV1 (KOS)
38	20	100.0	3936	6 E03026	E03026 DNA encodin
39	20	100.0	3996	6 E03114	E03114 DNA encodin
40	20	100.0	3997	6 E00357	E00357 DNA encodin
41	20	100.0	6098	14 HS1GBICP	M14164 HSV1 (strai
42	20	100.0	9756	14 HS1GLYB	M21629 Herpes simp
43	20	100.0	152261	14 H81CG	X14112 Human herpe
44	20	100.0	154746	14 HSV2RG52	Z66099 Herpes simp
45	18.4	92.0	1521	8 AF148506	AF148506 P15um bat

## ALIGNMENTS

RESULT 1  
LOCUS 103657  
DEFINITION Sequence 5 from Patent US 4642333.  
ACCESSION 103657  
VERSION 103657.1 GI:268613  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 798)  
PERSON Person, S.  
AUTHORS Immunologically reactive non-glycosylated amino acid chains of  
TITLE glycoprotein B of herpes virus types 1 and 2  
JOURNAL Patent: US 4642333-A 5 10-FEB-1987;

```

FEATURES
  source          600 Locust Ln.; State College, PA
                  1..798
                  /organism="unknown"
BASE COUNT      155 a 261 c 254 g 128 t
ORIGIN
Query Match          100.0%; Score 20; DB 6; Length 798;
Beat Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCACCACCGTCAGCACCTTC 20
        |||||||
        74 TCACCACCGTCAGCACCTTC 93

RESULT 2
S74390/c          1478 bp      DNA      VRL 30-APR-1995
DEFINITION      gBpath/UL27-glycoprotein B [herpes simplex virus type 1 HSV-1,
ACCESSION      S74390
VERSION      S74390
KEYWORDS      S74390.1 GI:786563
SOURCE
ORGANISM      Human herpesvirus 1
                Human herpesvirus 1
                Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
                Alphaherpesvirinae; Simplexvirus.
                1 (bases 1 to 1478)
                Koetal,M., Backl,I., Rajcani,J. and Kaerner,H.C.
                Replacement of glycoprotein B gene in the herpes simplex virus type
                1 strain ANGpath DNA by that originating from nonpathogenic strain
                KOS reduces the pathogenicity of recombinant virus
                Acta Virol. 38 (2), 77-88 (1994)
                95067449
                MEDLINE
                PUBMED
                7976866
REMARK      GenBank staff at the National Library of Medicine created this
                entry [NCBI gi2849158378] from the original journal article.
                This sequence comes from Fig. 8.
                Location/Qualifiers
                source          1..1478
                                /organism="Human herpesvirus 1"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:10298"
                                1..1478
                                /gene
                                /partial
                                /gene="gBpath/UL27"
                                /note="glycoprotein B"
BASE COUNT      236 a 451 c 528 g 263 t
ORIGIN
Query Match          100.0%; Score 20; DB 14; Length 1478;
Beat Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCACCACCGTCAGCACCTTC 20
        |||||||
        456 TCACCACCGTCAGCACCTTC 437

RESULT 3
S65444          1488 bp      DNA      VRL 21-APR-2003
LOCUS      S65444
DEFINITION      UL27-glycoprotein B [herpes simplex virus type 1 HSV-1, ANG,
ACCESSION      S65444
VERSION      S65444
KEYWORDS      S65444.1 GI:415494
SOURCE
ORGANISM      Human herpesvirus 1
                Human herpesvirus 1
                Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
                Alphaherpesvirinae; Simplexvirus.
                1 (bases 1 to 1488)
REFERENCE
1 (bases 1 to 1488)

```

**AUTHORS** Vihanez,S.A. and Stevens,J.G.  
**TITLE** Glycoprotein B is a specific determinant of herpes simplex virus type 1 neuroinvasiveness  
**JOURNAL MEDLINE** J.Virology. 67 (10), 5948-5954 (1993)  
**PUBMED** 8396662  
**REMARK** GenBank staff at the National Library of Medicine created this entry [NCBI gibbs 137392] from the original journal article.  
**FEATURES** This sequence comes from Fig. 7.  
    **source**  
        1..1488  
            `/organism="Human herpesvirus 1"`  
            `/mol_type="genomic DNA"`  
            `/db_xref="taxon:10298"`  
            <1..>1488  
            `/gene="UL27"`  
            <1..>1488  
            `/gene="UL27"`  
            `/note="This sequence comes from Fig. 7"`  
            `/codon_start=1`  
            `/product="glycoprotein B"`  
            `/protein_id="A01396.1"`  
            `/db_xref="GI:30027723"`  
            `/translation="VDFGVARDLTITKARAPPTTNRLTTPKFTVAMDWPKRPSVCT  
MTKKQVIDENLRSEHYGSFRPSSDAISTTTTLTLETRPLRVLDGDTGDADAMDR  
IFARRIVATIKVGQEPQYLVANGPLAIYOPLLLSNTLAEIYVEHLKEQSRRPNPTF  
PPKASANNASVERI KTTSISEPARLOPTYNHIQRHVDMDGRVAIAWCEIQNHETLM  
NEARKSNPAIASATIGVRVSARMELGVNAVSTCVPAADRNVIONSMRISIRGACY  
SRPLSVRFREDQPLVEGQLGENNELTDTDAIEPCVGHRRVFTEGGYVFEFYAY  
SHOISRDIITTFEPTDINLTMLDEHFLEVTREIRIDSGLDVTEVORNOIHDH  
LRPFADITVIHADANAMAFAGLGAFPFGMGDLGRAVGKVMGTIVGVSAVGSVSPM  
SNPGALAVGLVLVAGLAAPFAFRIVMRILQSNPMKALYLPTTKEL"`

**BASE COUNT** 260 A 530 C 459 G 239 T  
**ORIGIN**

**Query Match** 100.0%; Score 20; DB 14; Length 1488;  
**Best Local Similarity** 100.0%; Pred. No. 3,4e+02;  
**Matches** 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1028 TCACCACCGTCAGCACCTTC 1047  
Qy 1 TCACCACCGTCAGCACCTTC 20  

**RESULT 4**  
**LOCUS** HSiGBM 1857 bp DNA linear VRL 02-AUG-1993  
**DEFINITION** HSV1 (mutant strain tsB5), glycoprotein B (gB) gene.  
**ACCESSION** K02720  
**VERSION** K02720.1 GI:330087  
**KEYWORDS** glycoprotein.  
**SOURCE** Human herpesvirus 1  
**ORGANISM** Human herpesvirus 1  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
1 (bases 1 to 1857)  
Baik D.J., Fox B.A., Deluca N.A. and Person S.  
Nucleotide sequence of a region of the herpes simplex virus type 1 gB glycoprotein gene: mutations affecting rate of virus entry and cell fusion  
Virology 137 (1), 185-190 (1984)

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**COMMENT** Original source text: HSV1 strain tsB5 DNA, clone pTBG-B51.  
HSV1 strain tsB5 is a mutant of strain HFEM that is temperature dependent for virus growth, production of gB, and fusion of Vero and HEL cells. The syn and ts lesions of tsB5 are separable by recombination, the syn defect being located between 0.345 and 0.355 map units, and the ts defect being located between 0.360 and 0.368 map units. The rate of entry determinant of tsB5 is genetically separable from both the syn and ts lesions and is located between 0.350 and 0.360 map units, which is entirely within the glycoprotein B gene (0.346 to 0.366 map units).

Strain ts85 has a total of 12 base substitutions within the coding region for gB when compared to strain KOS. Of these, only two cause amino acid substitutions; a 't' to 'c' transition at bp 695 causes a Val to Ala substitution, and a 'g' to 'a' transition at bp 1610 causes an Arg to His substitution. Because the rate-of-entry determinant has been mapped to between 0.350 and 0.360 map units, the transition at bp 695 must be responsible for the difference in rate-of-entry between ts85 and KOS.

## FEATURES

## source

## CDS

```

1..1857
/organism="Human herpesvirus 1"
/mol_type="genomic DNA"
/db_xref="taxon:10298"
<1..1752
/note="glycoprotein B"
/codon_start=1
/protein_id="AAA45777.1"
/db_xref="GI:330088"
/translation="VDGFYARDLTKARATADPTTNLLTPKFTVAMDWPKRPSVCT
MTKQEVDEMISEVGSFRSSDAISTFTNTLSEYPLSRDLGDCIGKADAMDRI
IFARYNATHIKVGOQYVLANGFLAIAPLSTNLALVYRHLREGRKRPPTP
PPGASANAVERIKTTSIFBAPLOFTYNNHIOHYNDMLGRVATAMCLONHETLM
NEARKLNPAISATVGRVSRVSMGLDVAVSTCVPAVDNYIVONSRISSRPACV
SRPLVSRFEDGPLEEGQLGENNELRLTRDAIEPCTVGHRRYFTFGGIVYFESYAY
SHQLSRADITTVSTFDLNTLWLEDEHEFVPLEVYRHYKOSGLDYTEVORNLHD
LRFADIDIVIHADANAMAFAGAFEEGMDGAPVAGKVMGIVGVSAVGSVSG
SNPFALAVGLIVLALAAFPAPFVMELOGNPKMALPLTKELKYPNTNDASGEG
EEGGDFDEAKLAEAREMTRYMALVSAMETHEKAKKGSRLSKVTDVMWRKRRNT
NTQVYPNKDGDMDDEDL"

```

BASE COUNT 358 a 624 c 576 g 299 t  
 ORIGIN 1 bp upstream of SalI site; map coordinate 0.360.

Query Match 100.0%; Score 20; DB 14; Length 1857;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
 |||||  
 Db 1028 TCACCACCGTCAGCACCTTC 1047

## RESULT 5

E03093 2088 bp DNA linear PAT 29-SEP-1997  
 LOCUS E03093  
 DEFINITION DNA encoding truncated gB of herpes simplex virus typel (HSV-1).  
 ACCESSION E03093  
 VERSION E03093.1 GI:2171311  
 KEYWORDS JP 1991218397-A/2.  
 SOURCE Human herpesvirus 1  
 ORGANISM Human herpesvirus 1  
 Viruses: dadNA viruses, no RNA stage; Herpesviridae;  
 Alphaherpesvirinae; Simplexvirus.  
 1 (bases 1 to 2088)  
 Fujisawa,Y., Hinuma,K., Asakawa,N. and Otsuka,S.  
 POLYPEPTIDE AND ITS PRODUCTION  
 Patent: JP 1991218397-A 2 25-SEP-1991;  
 TAKEDA CHEM IND LTD  
 OS Herpes simplex virus typel (HSV-1)  
 PN JP 1991218397-A/2

## COMMENT

PC 25-SEP-1990 JP 1990161448  
 PR 22-JUN-1989 JP 89P 158238, 30-NOV-1989 JP 89P 308941 PI  
 FUJISAWA YUKIO, HINUMA KUNIJU, ASAKAWA NAKO, OTSUKA SACHIO PC  
 C07K13/00 C12N1/19, C12N15/38, C12P21/02//A61K39/245, (C12N1/19, PC  
 C12N1:865),  
 PC (C12P21/02, C12N1:865);  
 CC strandedness: Double;  
 CC topology: Linear;  
 CC hypothetical: No;  
 CC anti-sense: No;  
 CC \*source: strain=Miyama;  
 CC \*source: clone=PHS106 delta Tch;  
 FH Key Location/Qualifiers

FH CDS 1..2085  
 FT /product='truncated gB of HSV-1' FT 3'UTR  
 FT 2086..2088  
 FT mat\_peptide 1..2082  
 FT /product='truncated gB of HSV-1'.

## FEATURES

## source

## BASE COUNT

417 a 751 c 620 g 300 t

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2088;  
 Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
 |||||  
 Db 1904 TCACCACCGTCAGCACCTTC 1923

## RESULT 6

E00358 2586 bp DNA linear PAT 29-SEP-1997  
 LOCUS E00358  
 DEFINITION DNA encoding glycoprotein B of herpes simplex virus type 2.  
 ACCESSION E00358  
 VERSION E00358.1 GI:2168645  
 KEYWORDS JP 1985115529-A/2.  
 SOURCE unidentified  
 ORGANISM unidentified  
 unclassified.  
 1 (bases 1 to 2586)  
 Sutarre,P.  
 NON-GLYCOSTYLATED AMINO ACID CHAIN IMMUNOLOGICALLY REACTIVE TO  
 GLYCOPROTEIN OF HERPES VIRUS 1 AND 2  
 Patent: JP 1985115529-A 2 22-JUN-1985;  
 SUTARRE PAASON  
 OS herpes simplex virus type 2  
 PN JP 1985115529-A/2  
 PD 22-JUN-1985  
 PF 23-JUN-1984 JP 1984122915  
 PR 23-JUN-1983 US 83 506386, 16-SEP-1983 US 83 532996 PI  
 SUTARRE PAASON

## COMMENT

PC A61K39/245, C07H21/04, C07K15/14, C12N15/00//C12P21/02; CC  
 strandedness: Double;  
 CC topology: Linear;  
 CC hypothetical: No;  
 CC anti-sense: No;  
 CC \*source: strain=HG 52;  
 CC \*source: clone=p52BX;  
 CC \*source: map\_position=0.386-0.348;  
 FH Key Location/Qualifiers  
 FT 1..2727  
 FT /product='glycoprotein B of herpes simplex FT  
 FT virus type 2'.

## FEATURES

## source

## BASE COUNT

481 a 886 c 841 g 378 t

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2586;  
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
 |||||  
 Db 1917 TCACCACCGTCAGCACCTTC 1936

RESULT 7  
LOCUS BD145144 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Herpes simplex virus VP16 vaccine.  
ACCESSION BD145144.1 GI:27850902  
VERSION JP 2002136297-A/4.  
KEYWORDS unclassified  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2712)  
AUTHORS Burke, R.L. and Sekulovich, R.E.  
TITLES Herpes simplex virus VP16 vaccine  
JOURNAL Patent: JP 2002136297-A 4 14-MAY-2002;  
CHIRON CORP

COMMENT  
OS HSV gB1  
PN JP 2002136297-A/4  
PD 14-MAY-2002  
PR 10-SEP-2001 JP 2001277435  
PR 02-AUG-1990 US 561,528  
PI RAE LYN BURKE, ROSE E SEKULOVICH  
PC C12N15/09, A61K38/00, A61K39/245, A61P31/12, C07K14/03, C12N15/00,  
PC A61K37/02  
CC Herpes simplex virus VP16 vaccine  
FH Key Location/Qualifiers  
FT CDS (1)..(2712).

FEATURES  
source  
1..2712  
/organism="unclassified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

BASE COUNT 513 a 949 c 853 g 397 t

ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2712;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20  
|||||  
1988 TCACCAACCTGACGACCTTC 2007

Db

RESULT 8  
LOCUS BD165793 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Vaccine for use in the therapeutic treatment of HSV.  
ACCESSION BD165793  
VERSION BD165793.1 GI:27871605  
KEYWORDS JP 2002167398-A/2.  
SOURCE Herpes simplex virus 7  
ORGANISM Herpes simplex virus 7  
Virus; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE 1 (bases 1 to 2712)  
AUTHORS Burke, R.L., Pachl, C. and Valenzuela, P.D.T.  
TITLES Vaccine for use in the therapeutic treatment of HSV  
JOURNAL Patent: JP 2002167398-A 2 11-JUN-2002;  
CHIRON CORP

COMMENT  
OS Herpes simplex virus 7  
PN JP 2002167398-A/2  
PD 11-JUN-2002  
PR 12-SEP-2001 JP 2001277340  
PR 20-OCT-1986 US 921,213  
PI RAE LYN BURKE, CAROL PACHL, PABLO D T VALENZUELA PC  
C07K14/035, A61K39/245, A61P31/22, C12N15/09, C12N15/09, C12N15/00, PC  
C12N15/00  
CC Vaccine for use in the therapeutic treatment of HSV FH Key  
Location/Qualifiers  
FT CDS (1)..(2712).

FEATURES  
source  
1..(2712).

source 1..2712  
/organism="Herpes simplex virus 7"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:57277"

BASE COUNT 490 a 938 c 884 g 400 t

ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2712;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20  
|||||  
1982 TCACCAACCTGACGACCTTC 2001

Db

RESULT 9  
LOCUS BD165795 2712 bp DNA linear PAT 17-JAN-2003  
DEFINITION Vaccine for use in the therapeutic treatment of HSV.  
ACCESSION BD165795  
VERSION BD165795.1 GI:27871607  
KEYWORDS JP 2002167398-A/4.  
SOURCE Herpes simplex virus 7  
ORGANISM Herpes simplex virus 7  
Virus; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE 1 (bases 1 to 2712)  
AUTHORS Burke, R.L., Pachl, C. and Valenzuela, P.D.T.  
TITLES Vaccine for use in the therapeutic treatment of HSV  
JOURNAL Patent: JP 2002167398-A 4 11-JUN-2002;  
CHIRON CORP

COMMENT  
OS Herpes simplex virus 7  
PN JP 2002167398-A/4  
PD 11-JUN-2002  
PR 12-SEP-2001 JP 2001277340  
PR 20-OCT-1986 US 921,213  
PI RAE LYN BURKE, CAROL PACHL, PABLO D T VALENZUELA PC  
C07K14/035, A61K39/245, A61P31/22, C12N15/09, C12N15/09, C12N15/00, PC  
C12N15/00  
CC Vaccine for use in the therapeutic treatment of HSV FH Key  
Location/Qualifiers  
FT CDS (1)..(2712).

FEATURES  
source  
1..2712  
/organism="Herpes simplex virus 7"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:57277"

BASE COUNT 514 a 949 c 853 g 396 t

ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2712;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20  
|||||  
1991 TCACCAACCTGACGACCTTC 2010

Db

RESULT 10  
LOCUS BD145143 2715 bp DNA linear PAT 17-JAN-2003  
DEFINITION Herpes simplex virus VP16 vaccine.  
ACCESSION BD145143  
VERSION BD145143.1 GI:27850901  
KEYWORDS JP 2002136297-A/3.  
SOURCE unclassified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2715)  
AUTHORS Burke, R.L. and Sekulovich, R.E.  
TITLES Herpes simplex virus VP16 vaccine



JOURNAL Patent: JP 2002136297-A 3 14-MAY-2002;  
COMMENT CHIRON CORP  
OS HSV 9B2  
PN JP 2002136297-A/3  
PD 14-MAY-2002  
PF 10-SEP-2001 JP 2001274335  
PR 02-AUG-1990 US 561,528  
PI RAE LYN BURKE, ROSE B SEKULOVICH  
PC C12N15/09, A61K39/00, A61K39/245, A61P31/12, C07K14/03, C12N15/00,  
PC A61K37/02  
CC Herpes simplex virus VP16 vaccine  
FH key  
FT CDS  
FEATURES  
source  
1..2715  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"  
BASE COUNT 491 a 938 c 886 g 400 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TCACCACCGTCGACACCTTC 20  
Db 1982 TCACCACCGTCGACACCTTC 2001  
RESULT 11  
AF097023 2715 bp DNA linear VRL 07-APR-2000  
LOCUS Human herpesvirus 1 strain HSZP glycoprotein B (UL27) gene,  
DEFINITION complete cds.  
ACCESSION AF097023  
VERSION AF097023.1 GI:6165611  
KEYWORDS  
SOURCE  
ORGANISM  
Human herpesvirus 1  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE  
AUTHORS Kosovsky, J., Vojvodova, A., Oravcova, I., Kudelova, M., Matls, J. and Rajcani, J.  
TITLE Herpes simplex virus 1 (HSV-1) strain HSZP glycoprotein B gene: comparison of mutations among strains differing in virulence  
JOURNAL Virus Genes 20 (1), 27-33 (2000)  
MEDLINE 20227371  
PUBMED 1076304  
REFERENCE 2 (bases 1 to 2715)  
AUTHORS Kosovev, J., Vojvodova, A., Oravcova, I., Majorova, J., Kudelova, M. and Rajcani, J.  
TITLE Direct Submission  
JOURNAL Submitted (02-OCT-1998) Institute of Virology, Slovak Academy of Sciences, Dubravska cesta 9, Bratislava 842 46, Slovakia  
FEATURES  
source  
1..2715  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="HSZP"  
/db\_xref="taxon:10298"  
1..2715  
/gene="UL27"  
1..2715  
/gene="UL27"  
/function="essential for fusion between virion envelope and cell membrane"  
/note="envelope glycoprotein, gB"  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AAF04615.1"  
/db\_xref="GI:6165612"

/translation="NRGAPARGCMFVVMALLGLTVLVAAPSSSPGTGVAAT  
QANAGPATPAPALGAAPTGDPKPKKKKKPKNPPPPAGDNATVAAGHATLREHL  
DKASTANFVCPPTGATVVOEPORPCPTREGQNTGIVAVPEATLAPYEFK  
ATMYKDVTSQVWFGHGYVSOFGI PEDRAAPREVIDIKIAKVCSTACVAVENL  
ETTAHRDDEHDELMELKPAVATRTSRGHTTDLKYNPSRVEAFHRYGTTVCYBEV  
DARSVPYDEFLATGDPVMSPFYREGSHTEHTSYAADPFKQVDGVADLTTPKA  
RATAPATRLTLTPKFTVAMDVPYRPSVCTKQOEDEMRLSEVGSFRPSDAIS  
TFTTNLEVEPLSRVDLDCIGDKARDAMDRIFARVYATHIKVQPOYLANGLFI  
AYQPLSTLAEIYVREHREOSRDPNPPPPASANAVERIKTTSISEPARLOF  
TYNHQRYVNDMLGRVATACELQNHETLNMBAKLNPNALASATVGRVARNHGD  
VNAVSTCVPAADNVIVONSKRISRPACISPLVSEFYEQGPLVEQLBENNELR  
LTRDAIECTVGRHRYFTFGGYVYFEBYASGHQSRADITVVSFTIDNTMLBEH  
FVPLVATYRHEIKOSGLDYTEVQRNQIHDRFADIDTVIHADANAAMPAGIAPFE  
GNGDGRVAVGVIQGVASVAVSGFMSNPGALVAGLVLAGLAAPFAPRV  
MLOSNPMKALYPLTTKELNPTNPDSAGEGSGDPPBAKIAEREMIRVMAVSAM  
EHTERAKKKKGSALSLAKYTDVMVRKRNNTYTVQPKDQADEDDL"  
BASE COUNT 516 a 943 c 856 g 400 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 TCACCACCGTCGACACCTTC 20  
Db 1991 TCACCACCGTCGACACCTTC 2010  
RESULT 12  
AF295528 2715 bp mRNA linear VRL 01-SEP-2001  
LOCUS Human herpesvirus 2 glycoprotein B2 (gB2) mRNA, complete cds.  
DEFINITION AF295528  
ACCESSION AF295528  
VERSION AF295528.1 GI:15428575  
KEYWORDS  
SOURCE  
ORGANISM  
Human herpesvirus 2  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE  
AUTHORS Lee, H.H.  
TITLE Sequencing and high level expression of the HSV-2 strain G  
JOURNAL glycoprotein B gene using a baculovirus HcNPV  
REFERENCE 2 (bases 1 to 2715)  
AUTHORS Lee, H.H.  
TITLE Direct Submission  
JOURNAL Submitted (12-AUG-2000) Department of Biology, Konkuk University, 1  
Hwayang-dong, Kwangjin-gu, Seoul 143-701, Korea  
FEATURES  
source  
1..2715  
/organism="Human herpesvirus 2"  
/mol\_type="mRNA"  
/strain="G"  
/db\_xref="taxon:10310"  
1..2715  
/gene="gB2"  
1..2715  
/gene="gB2"  
/codon\_start=1  
/product="glycoprotein B2"  
/protein\_id="AAK97852.1"  
/db\_xref="GI:15428576"  
/translation="NRGGGLICALVYGLVAVAASAPAAPAPASGCVATVAANG  
GPASRPVPSPTATTKARKKPKPPEATPPPDANAIVAAGHATLRAHLEIKVE  
NADAPVYCPPTGATVVOFQPRPCPTREGQNTGIVAVPEATLAPYEFKATMY  
KDVTSQVWFGHYSQFMGIFEDRAVPEVEYIDIKINAKVGRSTAKYVNMETAF  
KDHDETHMELKPAVATRTSRGHTTDLKYNPSRVEAFHRYGTTVCYBEVARSV  
YVBEFLATGDPVMSPFYREGSHTEHTSYAADRFKQVDGFARBDLTAKATSP  
TTRNLITPKFTVAMDVPYRPSVCTKQOEDEMRLSEVGSFRPSDAISTFTT  
NLTVSLSRVDLDCIGDKARDARAI DMRARKNATHIKVGOYVLAGGLIATOP  
LSNTLAEIYVREHREOSRDPNPPPPASANAVERIKTTSISEPARLOFTYN  
HQRVNDMLGRVATACELQNHETLNMBAKLNPNALASATVGRVARSAMLDGVA

Query Match 100.0%; Score 20; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCGTCAGCACTTC 20  
|||||  
1982 TCACCAACCGTCAGCACTTC 2001

RESULT 13  
AF311740 2715 bp DNA linear VRL 24-JAN-2001  
LOCUS Human herpesvirus 1 strain KOS glycoprotein B gene, complete cds.  
ACCESSION AF311740  
VERSION AF311740.1 GI:11321322  
KEYWORDS  
SOURCE Human herpesvirus 1  
ORGANISM Human herpesvirus 1  
Virusess; deDNA virusess, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE 1 (bases 1 to 2715)  
AUTHORS Percel, P.E., Fridberg, A., Parish, M.L. and Spear, P.G.  
TITLE Cell fusion induced by herpes simplex virus glycoproteins gB, gD, and gH-gL requires a gD receptor but not necessarily heparan sulfate  
JOURNAL Virology 279 (1), 313-324 (2001)  
MEDLINE 21066717  
PUBMED 11145912

REFERENCE 2 (bases 1 to 2715)  
AUTHORS Percel, P.E., Fridberg, A., Parish, M.L. and Spear, P.G.  
TITLE Direct Submision  
JOURNAL Submitted (05-OCT-2000) Microbiology-Immunology, Northwestern University Medical School, 320 E. Superior Ave., Ward 6-241, Chicago, IL 60611, USA  
Location/Qualifiers  
1..2715  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="KOS"  
/db\_xref="taxon:10298"  
1..2715  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AA34116.1"  
/db\_xref="GI:11321323"  
/translation="MGQAPARGRRMFVWVALLGLTGLVLSAAPSFGTGGVAAAT  
QAANGPATPAPAPGAPGDPKPKKPKKPPPPRAGDNATVAAGHATLREHLR  
DIKANTDANFVCPPTGATVVOFOPRCPPTREGONTEGIAVFEATLAPYKFK  
ATMYKXDVTSQVWFGHRSQFMGIFEDBAPVPEFVIDIKNAKGCSTAKYVRNL  
ETTAHRDDEHETMELKAPANAATRSRGHTTDLKPNRSVAFPRYCTTACIVSEV  
DARSVPYVEFLATGDFVYMSPGYRGSGHTHTSTADRFKQDGFYAADLTTKA  
RATAPTRNLTTPKFTVAMDVWPKRPSVCTMKQEVEMLRSEYGSFRSSDLS  
TFTTNLTVEPLSRVLDGDCIGKADANDRIIPARYNATHIKVQPOYLANGFLLI  
AYOPLSNTLAEIYVREHLREBSRKPPNTPPPGASANAVERIKTSSIFARLOF  
TYNHIGHVNDMLGRVALANCELOHNEHLTMEARKLNPAIATSVYGRVSARMYGD  
VMASTCVPAADNVITVQNSMRISRSRGACYSRPLVSFRYEDGQPLVEQGLNNELR  
LTRDAIEPCTVHRRYFTFGGIVYFEEYASHQSRADITVSTFTIDNTMLEBHE  
FVLEVYTHHEIKDSGLDYTEVORNOHLDRFADIDTVIHADANAAMFAGAFVE  
GMGDLRAVGKVMGIVGVSAVGSFSMNPFGALAVGLVLAGLAAMFAGAFVE  
MLOSNPMKALYPLTTKELKNPTNPDSAGEGDFDAKLAEMEMRYMALVSAM  
ERTEHAKKKGTGTSALLSAKVTDMWAKRNTNTYTVPNKDGDDEDDL"

BASE COUNT 521 a 948 c 849 g 397 t

Query Match 100.0%; Score 20; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCGTCAGCACTTC 20  
|||||  
1991 TCACCAACCGTCAGCACTTC 2010

RESULT 14  
HH049121 2715 bp DNA linear VRL 08-MAR-1996  
LOCUS Human herpesvirus 1 strain ANG path, glycoprotein B (gB) gene, complete cds.  
ACCESSION HH049121  
VERSION U49121.1 GI:1218049  
KEYWORDS  
SOURCE Human herpesvirus 1  
ORGANISM Human herpesvirus 1  
Virusess; deDNA virusess, no RNA stage; Herpesviridae;  
Alphaherpesvirinae; Simplexvirus.  
REFERENCE 1 (bases 1 to 2715)  
AUTHORS Holland, T.C. and Saharikhiz-Langroodi, A.  
TITLE Direct Submision  
JOURNAL Submitted (13-FEB-1996) Dept. of Immunology and Microbiology, Wayne State University, 540 E. Canfield Ave., Detroit, MI 48201, USA  
Location/Qualifiers  
1..2715  
/organism="Human herpesvirus 1"  
/mol\_type="genomic DNA"  
/strain="ANG path"  
/db\_xref="taxon:10298"  
1..2715  
/gene="gB"  
/note="HSV-1 UL27 open reading frame"  
1..2715  
/gene="gB"  
/note="envelope glycoprotein"  
/codon\_start=1  
/product="glycoprotein B"  
/protein\_id="AA091805.1"  
/db\_xref="GI:1218050"  
/translation="MGQAPARGRRMFVWVALLGLTGLVLSAAPSFGTGGVAAAT  
QAANGPATPAPAPGAPGDPKPKKPKKPPPPRAGDNATVAAGHATLREHLR  
DIKANTDANFVCPPTGATVVOFOPRCPPTREGONTEGIAVFEATLAPYKFK  
ATMYKXDVTSQVWFGHRSQFMGIFEDBAPVPEFVIDIKNAKGCSTAKYVRNL  
ETTAHRDDEHETMELKAPANAATRSRGHTTDLKPNRSVAFPRYCTTACIVSEV  
DARSVPYVEFLATGDFVYMSPGYRGSGHTHTSTADRFKQDGFYAADLTTKA  
RATAPTRNLTTPKFTVAMDVWPKRPSVCTMKQEVEMLRSEYGSFRSSDLS  
TFTTNLTVEPLSRVLDGDCIGKADANDRIIPARYNATHIKVQPOYLANGFLLI  
AYOPLSNTLAEIYVREHLREBSRKPPNTPPPGASANAVERIKTSSIFARLOF  
TYNHIGHVNDMLGRVALANCELOHNEHLTMEARKLNPAIATSVYGRVSARMYGD  
VMASTCVPAADNVITVQNSMRISRSRGACYSRPLVSFRYEDGQPLVEQGLNNELR  
LTRDAIEPCTVHRRYFTFGGIVYFEEYASHQSRADITVSTFTIDNTMLEBHE  
FVLEVYTHHEIKDSGLDYTEVORNOHLDRFADIDTVIHADANAAMFAGAFVE  
GMGDLRAVGKVMGIVGVSAVGSFSMNPFGALAVGLVLAGLAAMFAGAFVE  
MLOSNPMKALYPLTTKELKNPTNPDSAGEGDFDAKLAEMEMRYMALVSAM  
ERTEHAKKKGTGTSALLSAKVTDMWAKRNTNTYTVPNKDGDDEDDL"

BASE COUNT 513 a 951 c 854 g 397 t

Query Match 100.0%; Score 20; DB 14; Length 2715;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCGTCAGCACTTC 20  
|||||  
1991 TCACCAACCGTCAGCACTTC 2010

RESULT 15  
AF259899 2809 bp DNA linear VRL 16-MAY-2000  
LOCUS



**BEST AVAILABLE COPY**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 17:02:34 ; Search time 100.198 Seconds  
(without alignments)  
538.821 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20  
Sequence: 1 tcaccaccgtcagcacccttc 20

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

N\_Geneseq\_13jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	2088	12 AAQ14479	Truncated HSVB ge
2	20	100.0	2645	19 AAV62158	HSV-2 strain SB5 C
3	20	100.0	2712	8 AAN71303	Herpes Simplex Vir
4	20	100.0	2724	8 AAN71399	Herpes Simplex Vir
5	20	100.0	2727	6 AAN50517	Sequence encoding
6	20	100.0	2841	19 AAV62146	HSV-2 strain SB5 C
7	20	100.0	2925	12 AAQ14455	HSV surface antigen
8	20	100.0	3098	7 AAN60195	Herpes simplex vir

9	20	100.0	3461	8 AAN70681	Recombinant herpes
10	20	100.0	3465	12 AAQ14478	HSVB gene. Herpe
11	20	100.0	3471	18 AAT93650	Herpes simplex vir
12	20	100.0	3472	9 AAN80907	Sequence of Herpes
13	20	100.0	3472	14 AAQ48496	Glycoprotein B (gB
14	20	100.0	3472	14 AAQ48497	Glycoprotein B (gB
15	20	100.0	3473	18 AAT93651	Herpes simplex vir
16	20	100.0	3474	9 AAN80908	Sequence of Herpes
17	20	100.0	3474	9 AAN80908	Sequence of Herpes
18	20	100.0	3474	9 AAN80908	Sequence of Herpes
19	20	100.0	3474	9 AAN80908	Sequence of Herpes
20	20	100.0	3474	9 AAN80908	Sequence of Herpes
21	20	100.0	3474	9 AAN80908	Sequence of Herpes
22	20	100.0	3474	9 AAN80908	Sequence of Herpes
23	20	100.0	3474	9 AAN80908	Sequence of Herpes
24	20	100.0	3474	9 AAN80908	Sequence of Herpes
25	20	100.0	3474	9 AAN80908	Sequence of Herpes
26	20	100.0	3474	9 AAN80908	Sequence of Herpes
27	20	100.0	3474	9 AAN80908	Sequence of Herpes
28	20	100.0	3474	9 AAN80908	Sequence of Herpes
29	20	100.0	3474	9 AAN80908	Sequence of Herpes
30	20	100.0	3474	9 AAN80908	Sequence of Herpes
31	20	100.0	3474	9 AAN80908	Sequence of Herpes
32	20	100.0	3474	9 AAN80908	Sequence of Herpes
33	20	100.0	3474	9 AAN80908	Sequence of Herpes
34	20	100.0	3474	9 AAN80908	Sequence of Herpes
35	20	100.0	3474	9 AAN80908	Sequence of Herpes
36	20	100.0	3474	9 AAN80908	Sequence of Herpes
37	20	100.0	3474	9 AAN80908	Sequence of Herpes
38	20	100.0	3474	9 AAN80908	Sequence of Herpes
39	20	100.0	3474	9 AAN80908	Sequence of Herpes
40	20	100.0	3474	9 AAN80908	Sequence of Herpes
41	20	100.0	3474	9 AAN80908	Sequence of Herpes
42	20	100.0	3474	9 AAN80908	Sequence of Herpes
43	20	100.0	3474	9 AAN80908	Sequence of Herpes
44	20	100.0	3474	9 AAN80908	Sequence of Herpes
45	20	100.0	3474	9 AAN80908	Sequence of Herpes

## ALIGNMENTS

AAQ14479	AAQ14479 standard; DNA; 2088 BP.
AC	AAQ14479;
DT	25-MAR-2003 (updated)
DT	23-JAN-1992 (first entry)
XX	Truncated HSVB gene.
XX	Vaccine; antigen; ss.
OS	Herpes simplex virus.
PH	Key
FT	Location/Qualifiers
FT	CDS 1..2088
XX	/*tag= a
PN	JP03218397-A.
PD	25-SEP-1991.
XX	
PF	21-JUN-1990; 90JP-0161448.
XX	
PR	30-NOV-1989; 89JP-0308941.
PR	22-JUN-1989; 89JP-0158236.
PR	21-JUN-1990; 90JP-0161448.
XX	
PA	(TAKE ) TAKEDA CHEM IND LTD.
XX	

Recombinant herpes  
HSVB gene. Herpe  
Herpes simplex vir  
Sequence of Herpes  
Glycoprotein B (gB  
Glycoprotein B (gB  
Herpes simplex vir  
Sequence of Herpes  
DNA fragment encod  
Sequence encoding  
HSV-2 strain SB5 C  
Human herpesvirus  
DNA encoding T. ch  
Triangulumvirine (T  
DNA encoding T. ch  
Complete nucleotid  
Plant D-like cycli  
Arabidopsis thalia  
Drosophila melanog  
Human foetal liver  
Probe #2089 for ge  
Human brain expres  
Human liver single  
Human liver single  
Human foetal liver  
Aspergillus oryzae  
DNA encoding novel  
Human ORF2 ORF2970  
Human olfactory re  
Human G protein co  
DNA encoding novel  
DNA encoding novel  
Human GCRC-33 CDN  
DNA encoding a hum  
DNA encoding a hum  
Aspergillus fumiga

DR WPI; 1991-328397/45.  
 DR P-PSDB; AAR14666.  
 XX  
 PT HSVB polypeptide(s) obt'd. by recombinant DNA techniques -  
 PT useful as vaccines against HSV and in diagnosis, can be produced  
 PT cheaply and safely.  
 XX  
 PS Claim 3; Fig 7; 24pp; Japanese.  
 CC The sequence encodes a truncated form of the HSVB polypeptide.  
 CC The recombinant protein can be used to prepare vaccines for  
 CC prophylaxis of HSV infection and for use in diagnostic kits.  
 CC See also AA014478.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 XX Sequence 2088 BP; 417 A; 751 C; 620 G; 300 T; 0 other;  
 SO

Query Match 100.0%; Score 20; DB 12; Length 2088;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGCACCCTTC 20  
 DB 1904 TCACCAACGTCAGCACCCTTC 1923

RESULT 2  
 AAV62158/C  
 ID AAV62158 standard; DNA; 2645 BP.  
 XX  
 AC AAV62158;  
 XX  
 DT 18-DEC-1998 (first entry)  
 XX  
 DE HSV-2 strain SBS Contig ID 14 DNA sequence.  
 XX  
 KW HSV-2 strain SBS; immunological response induction; therapy;  
 KW antiviral identification; viral protein inhibitor; ss.  
 XX  
 OS Herpes simplex virus type 2.  
 XX

Key Location/Qualifiers  
 FT CDS complement (81..2645)  
 FT /\*tag= a  
 FT /product= "ORF#1 protein"  
 XX  
 PN MO9820016-A1.  
 XX  
 PD 14-MAY-1998.  
 XX  
 PF 31-OCT-1997; 97MO-US20016.  
 XX  
 PR 09-JUN-1997; 97US-0049018.  
 PR 04-NOV-1996; 96US-0030279.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Chan JY, Dabrowski-Amaraal CE, Delvecchio AM, Dillon SB;  
 PI Esber KM, Leary JJ;  
 DR WPI; 1998-286847/25.  
 DR P-PSDB; AAW72113.  
 XX  
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
 PT and treatment of infection or inducing immunological response in  
 PT mammal  
 XX  
 PS Claim 1; Page 406-407; 748pp; English.  
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA  
 CC sequence of the invention. This sequence was isolated from HSV-2 strain  
 CC SBS (deposited as ATCC VR-2546), is designated Contig ID 14, and encodes  
 CC a HSV-2 protein. The protein can be used for the treatment or

CC prevention of disease, to induce an immunological response in a mammal or  
 CC to identify inhibitors, activators or novel antivirals. Antagonists of  
 CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence  
 CC or a vector containing it can also be used to induce an immunological  
 CC response in a mammal.  
 XX  
 SQ Sequence 2645 BP; 404 A; 837 C; 892 G; 512 T; 0 other;  
 SO

Query Match 100.0%; Score 20; DB 19; Length 2645;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGCACCCTTC 20  
 DB 814 TCACCAACGTCAGCACCCTTC 795

RESULT 3  
 AAN71303  
 ID AAN71303 standard; DNA; 2712 BP.  
 XX  
 AC AAN71303;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 30-APR-1991 (first entry)  
 XX  
 DE Herpes Simplex Virus-1 GB from PKBX.  
 XX  
 KW Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;  
 KW glycoprotein; GB; ss.  
 XX  
 OS Herpes simplex virus type 1 (KOS).  
 XX  
 PN US4642333-A.  
 XX  
 PD 10-FEB-1987.  
 XX  
 PF 20-JUN-1984; 84US-0622496.  
 XX  
 PR 20-JUN-1984; 84US-0622496.  
 XX  
 PA (PERS/) PERSON S.  
 XX  
 PI Person S;  
 XX  
 DR WPI; 1987-056354/08.  
 DR P-PSDB; AAW71135.  
 XX  
 PT Amino acid chain of glyco:protein B of HSV-1 and 2 - prep'd. as  
 PT recombinant and used for vaccines for herpes simplex virus types 1  
 PT and 2.  
 XX  
 PS Example; Table 2; 16pp; English.  
 CC A pure non-glycosylated amino acid (AA) chain comprising a sequence  
 CC corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2  
 CC virus which is antigenic to HSV-1 of HSV-2, which contains no more  
 CC than 750 AA residues, and which includes AA residues 135-649  
 CC inclusive is claimed. It can be used to produce vaccines for  
 CC prophylaxis and treatment of HSV-1 and HSV-2.  
 CC See AAN71399 for the HSV-2 sequence.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC  
 XX Sequence 2712 BP; 519 A; 945 C; 849 G; 397 T; 2 other;  
 SO

Query Match 100.0%; Score 20; DB 8; Length 2712;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGCACCCTTC 20  
 DB 1988 TCACCAACGTCAGCACCCTTC 2007

```

XX          RESULT 4
XX ID      AAN71399 standard; DNA; 2724 BP.
XX AC      AAN71399;
XX DT      25-MAR-2003 (updated)
XX DT      30-APR-1991 (first entry)
XX DE      Herpes Simplex Virus-2 gB from p52BXK.
XX KW      Vaccine; prophylaxis; treatment; Herpes Simplex Virus-1;
XX KW      glycoprotein; gB; ss.
XX OS      Herpes simplex virus type 2 HGS2.
XX PN      US464233-A.
XX PD      10-FEB-1987.
XX PE      20-JUN-1984;   84US-0622496.
XX PR      20-JUN-1984;   84US-0622496.
XX PA      (PERS/) PERSON S.
XX PI      Person S;
XX PS      WPI; 1987-056354/08.
XX DR      P-PDB; AAP71136.
XX PT      Amino acid chain of glyco:protein B of HSV-1 and 2 - prepd. as
XX FT      recombinant and used for vaccines for herpes simplex virus types 1
XX         and 2.
XX         Example; Table 2; 16pp; English.
XX PS      'N' indicated nucleotide to be determined (corresp. to residues
CC CC      17-41 of HSV-1 gB).
CC CC      A pure non-glycosylated amino acid (AA) chain comprising a sequence
CC CC      corresponding to that occurring in glycoprotein B of HSV-1 or HSV-2
CC CC      virus which is antigenic to HSV-1 of HSV-2, which contains no more
CC CC      than 750 AA residues, and which includes AA residues 135-649
CC CC      inclusive is claimed. It can be used to produce vaccines for
CC CC      prophylaxis and treatment of HSV-1 and HSV-2.
CC CC      See AAN71303 for the HSV-1 sequence.
CC CC      (Updated on 25-MAR-2003 to correct PR field.)
XX SQ      Sequence 2724 BP; 494 A; 911 C; 854 G; 390 T; 75 other;

Query Match           100.0%; Score 20; DB 8; Length 2724;
Best Local Similarity 100.0%; Pred. No. 24;
Matches    20; Conservative    0; Mismatches    0; Indels    0; Gaps    0.

Oy              1 TCACCACCGTCGAGACTTTC 20
                |||||
Db              1991 TCACCACCGTCGAGACTTTC 2010

RESULT 5
ID      AAN50517
XX      AAN50517 standard; DNA; 2727 BP.
XX AC      AAN50517;
XX DT      21-NOV-1991 (first entry)
XX DE      Sequence encoding the glycoprotein B (gB) of Herpes simplex virus
XX         type 2 (HSV-2).
XX KW      Vaccine; glycoprotein B; ss.

```

OS	Herpes simplex virus type 2.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers 1..2727 /*tag= a
XX	
PN	EPI33063-A.
XX	
PD	13-FEB-1985.
XX	
PF	22-JUN-1984; 84EP-0401312.
XX	
PR	16-SBP-1983; 83US-0532996.
PR	23-JUN-1983; 83US-0506986.
FR	20-JUN-1984; 84US-0622496.
PA	(PERS/) PERSON S.
XX	
PI	Person S;
XX	
DR	WPI; 1985-039636/07.
XX	
PT	New amino acid chains of glyco-protein B of herpes simplex virus
PT	- prepd. by recombination DNA methods for use in vaccines
XX	
PS	Disclosure; Table 2, Page 35-35C, 40pp, English.
XX	
CC	The inventors claim a non-glycosylated amino acid chain comprising a
CC	sequence corresponding to that occurring in glycoprotein B of HSV-1
CC	or HSV-2. Preferably, when the chain comprises a sequence corresp.
CC	to that occurring in glycoprotein B of HSV-2, it contains 4-750 AA
CC	residues. The chain esp. has a molecular wt. of 6500 daltons and
CC	includes AA residues 165-629 from the HSV-2 glycoprotein B sequence.
SQ	
	Sequence 2727 BP; 493 A; 913 C; 855 G; 391 T; 75 other;
QY	
	Query Match 100.0%; Score 20; DB 6; Length 2727; Best Local Similarity 100.0%; Pred. No. 24; Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	
	1 TCACCACCGTCAGCACCTTC 20                     1994 TCACCACCGTCAGCACCTTC 2013
RESULT 6	
AAV62146	
ID	AAV62146 standard; DNA; 2841 BP.
XX	
AC	AAV62146;
XX	
DT	07-DEC-1998 (first entry)
XX	
DE	HSV-2 strain SBS Contig ID 91 DNA sequence.
XX	
KW	HSV-2 strain SBS; immunological response induction; therapy;
KW	antiviral identification; viral protein inhibitor; ss.
XX	
OS	Herpes simplex virus type 2.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers 358..2745 /*tag= a
XX	
PN	WO9820016-A1.
XX	
PD	14-MAY-1998.
XX	
PF	31-OCT-1997; 97WO-US20016.
XX	
PR	09-JUN-1997; 97US-0049018.
FR	04-NOV-1996; 96US-0030279.
XX	

PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;  
 PI Esser KM, Leary JJ;  
 XX  
 DR WPI; 1998-286847/25.  
 DR P-PSDB; AAW72062.  
 XX  
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention  
 PT and treatment of infection or inducing immunological response in  
 PT mammal  
 XX  
 PS Claim 1; Page 284-285; 748pp; English.  
 XX  
 CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA  
 CC sequence of the invention. This sequence was isolated from HSV-2 strain  
 CC SBS (deposited as ATCC VR-2546), is designated Contig ID 91, and encodes  
 CC a HSV-2 protein. The proteins can be used for the treatment or  
 CC prevention of disease, to induce an immunological response in a mammal or  
 CC to identify inhibitors, activators or novel antivirals. Antagonists of  
 CC the protein can be used to inhibit a viral polypeptide. The DNA sequence  
 CC or a vector containing it can also be used to induce an immunological  
 CC response in a mammal.  
 XX  
 SQ Sequence 2841 BP; 577 A; 921 C; 915 G; 425 T; 3 other;  
 XX  
 Query Match 100.0%; Score 20; DB 19; Length 2841;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TCACCAACCGTCAGACCTTC 20  
 DB 2012 TCACCAACCGTCAGACCTTC 2031  
 XX  
 RESULT 7  
 ID AAQ14455  
 XX AAQ14455 standard; DNA; 2925 BP.  
 AC  
 XX AAQ14455;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 21-JAN-1992 (first entry)  
 XX  
 DE HSV surface antigens gB/gD fusion protein gene.  
 XX  
 KM Herpes simplex virus; vaccine; ss.  
 XX  
 OS Herpes simplex virus.  
 XX  
 PI Key Location/Qualifiers  
 PI CDS 1..2925  
 FT /\*tag= a  
 FT  
 XX  
 XX JF03220200-A.  
 PN  
 XX  
 PD 27-SEP-1991.  
 PD  
 XX  
 PF 29-NOV-1990; 90JP-0325474.  
 PF  
 XX  
 PR 30-NOV-1989; 89JP-0308942.  
 PR  
 XX 29-NOV-1990; 90JP-0325474.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 PA  
 DR WPI; 1991-329235/45.  
 DR  
 DR P-PSDB; AAR14680.  
 XX  
 XX Polypeptide for herpes simplex virus vaccine prodn. - obtd. by  
 PT culturing transformant of recombinant DNA coding fused protein of  
 PT virus surface antigen  
 XX  
 PS Claim 6; Fig 8; 16pp; Japanese.  
 }

XX  
 CC The DNA sequence encodes a fusion protein comprising herpes simplex  
 CC virus (HSV) surface antigens gB and gD. This fusion protein shows  
 CC the same physiological activity as surface antigen prepd. from HSV  
 CC infected cells. It can be used in a vaccine for the prevention of  
 CC HSV infection.  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 2925 BP; 579 A; 1061 C; 848 G; 437 T; 0 other;  
 XX  
 Query Match 100.0%; Score 20; DB 12; Length 2925;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 TCACCAACCGTCAGACCTTC 20  
 DB 1904 TCACCAACCGTCAGACCTTC 1923  
 XX  
 RESULT 8  
 ID AAN60195  
 XX AAN60195 standard; DNA; 3098 BP.  
 AC  
 XX AAN60195;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 31-JUL-1991 (first entry)  
 XX  
 DE Herpes simplex virus glycoprotein gB gene.  
 XX  
 KM HSV; gB glycoprotein; vaccine; ss.  
 XX  
 OS Herpes simplex virus.  
 XX  
 PI Key Location/Qualifiers  
 PI CDS 257..2966  
 FT /\*tag= a  
 FT  
 XX  
 PN EP170169-A.  
 PN  
 XX  
 PD 05-FEB-1986.  
 PD  
 XX  
 PF 19-JUL-1985; 85EP-0109042.  
 PF  
 XX  
 PR 20-JUL-1984; 84JP-0151766.  
 PR  
 XX 11-DEC-1984; 84JP-0262465.  
 XX  
 PA (KAGA ) CHEMO SERO THERAPEUTIC RES INST.  
 PA  
 PI Nozaki C, Makizumi K, Kino Y, Eto, T, Ohtomo N;  
 PI  
 XX  
 DR WPI; 1986-036935/06.  
 DR  
 DR P-PSDB; AAP60244.  
 XX  
 XX Recombinant DNA containing herpes simplex virus gene or fragment -  
 PT useful in transformant yeast for prodn. of high purity herpes  
 PT simplex virus glycoproteins for use in protective vaccines.  
 XX  
 PS Disclosure; Fig. 6; 53pp; English.  
 XX  
 CC The gene encodes a gB glycoprotein which is useful for the  
 CC production of vaccines conferring protection against herpes simplex  
 CC virus infections. The vaccine is safe in use and does not contain a  
 CC carcinogenic gene, as with prior vaccines.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 3098 BP; 586 A; 1070 C; 959 G; 483 T; 0 other;  
 XX  
 Query Match 100.0%; Score 20; DB 7; Length 3098;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX



Qy 1 TCACCAACGCTCAGACCTTC 20  
 |||||  
 Db 2244 TCACCAACGCTCAGACCTTC 2263

# RESULT 9

ID AAN70681 standard; DNA; 3461 BP.

AC AAN70681;

DT 25-MAR-2003 (updated)

DT 10-MAR-2003 (updated)

DT 30-APR-1991 (first entry)

DE Recombinant herpes simplex virus gene.

KM Herpes simplex virus; E.coli; plasmid; vector; promoter; antigen;

OS Herpes simplex virus.

FT CDS Key Location/Qualifiers

FT 254..2962 /\*tag= a

FT /product= HSV\_GB

PN JP62115288-A.

PD 26-MAY-1987.

PF 14-NOV-1985; 85JP-0255971.

PR 14-NOV-1985; 85JP-0255971.

PA (KAGA ) KAGAKU OYOBI KESSEI RYOH.

DR WPI; 1987-183377/26.

DR P-PSDB; AAP70426.

PT Recombinant DNA contg. herpes simplex virus gene - is integrated in

PT E.coli plasmid vector, contg. incorporated promoter region for

PT animal cells, giving antigen.

PS Dielosure; Fig 3(1-3); 13pp; Japanese.

CC The gene encoding HSV GB is cloned into E.coli. To increase the

CC efficiency of expression, the gene encoding the C-terminal region

CC of GB, which is very hydrophobic, is removed and the remaining gene

CC is cloned into an expression vector. The vector is composed of pBR322,

CC SV40 early promoter and the HSV GB gene.

CC The construct is useful for the production of a good antigen for

CC a HSV vaccine.

CC (updated on 10-MAR-2003 to add missing OS field.)

CC (updated on 25-MAR-2003 to correct PA field.)

CC

SQ Sequence 3461 BP; 666 A; 1188 C; 1041 G; 566 T; 0 other;

Query Match 100.0%; Score 20; DB 8; Length 3461;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACCAACGCTCAGACCTTC 20  
 |||||  
 Db 2244 TCACCAACGCTCAGACCTTC 2260

# RESULT 10

AAQ14478 standard; DNA; 3465 BP.

AC AAQ14478;

DT 25-MAR-2003 (updated)

DT 23-JAN-1992 (first entry)

DE HSVGB gene.

KM Vaccine; antigen; ss.

OS Herpes simplex virus.

FT CDS Key Location/Qualifiers

FT 251..2965 /\*tag= a

PN JP03218397-A.

PD 25-SEP-1991.

PF 21-JUN-1990; 90JP-0161448.

PR 30-NOV-1989; 89JP-0308941.

PR 22-JUN-1989; 89JP-0158238.

PR 21-JUN-1990; 90JP-0161448.

PA (TAKA ) TAKEDA CHEM IND LTD.

DR WPI; 1991-328397/45.

DR P-PSDB; AAR14665.

PT HSVGB polypeptide(s) obt'd. by recombinant DNA techniques -

PT useful as vaccines against HSV and in diagnosis, can be produced

PT cheaply and safely.

PS Claim 3; Fig 1; 24pp; Japanese.

CC The sequence encodes the HSVGB polypeptide. The recombinant protein

CC or esp. truncated forms of it, can be used to prepare vaccines for

CC prophylaxis of HSV infection and for use in diagnostic kits.

CC See also AAQ14479.

CC (updated on 25-MAR-2003 to correct PA field.)

CC

SQ Sequence 3465 BP; 666 A; 1194 C; 1044 G; 561 T; 0 other;

Query Match 100.0%; Score 20; DB 12; Length 3465;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACCAACGCTCAGACCTTC 20  
 |||||  
 Db 2244 TCACCAACGCTCAGACCTTC 2260

RESULT 11

AAQ93650 standard; DNA; 3471 BP.

AC AAQ93650;

DT 26-FEB-1998 (first entry)

DE Herpes simplex virus type 1 glycoprotein B encoding DNA.

KM Herpes simplex virus; HSV; type I; type II; glycoprotein B; vaccine;

KW membrane protein; virus-specific glycoprotein;

KX transmembrane anchor region; ss.

OS Herpes simplex virus type 1.

FT Key Location/Qualifiers

FT CAAT\_signal 55..62

FT /\*tag= a

FT /note= "putative"

FT TATA\_signal 125..131

FT /\*tag= b

```

FT      CDS                               /note="putative"
FT      309..3023
FT      /*tag= C
FT      polyA_signal                      /tag= d
FT      3166..3173
FT      /*tag= "putative"
FT      polyA_signal                      /note="putative"
FT      3409..3416
FT      /*tag= e
FT      /note= "putative"
XX      US5648079-A.
XX      PD
XX      15-JUL-1997.
XX      PF
XX      08-DEC-1994; 94US-0351875.
XX      PR
XX      20-OCT-1986; 86US-0921730.
XX      PR
XX      06-APR-1984; 84US-0597784.
XX      PR
XX      20-SEP-1990; 90US-0587179.
XX      PR
XX      21-DEC-1992; 92US-0993415.
XX      PR
XX      18-OCT-1993; 93US-0138717.
XX      PA
XX      (CHIR ) CHIRON CORP.
XX      PI
XX      Burke LR, Pachl C, Valenzuela PDT;
XX      WPI; 1997-372032/34.
XX      DR
XX      P-PSDB; AAM34552.
XX      PT
XX      Vaccines against herpes simplex virus infection - containing
XX      recombinant HSV glycoprotein B
XX      PS
XX      Disclosure; Fig 4A-F; 33pp; English.
XX      CC
XX      The present sequence is from the Herpes simplex virus (HSV) type I
XX      strain Patcon and encodes a glycoprotein B. HSV is a double stranded
XX      CC virus packaged within an icosahedral nucleocapsid enveloped within a
XX      CC membrane. The membrane includes a number of virus-specific
XX      CC glycoproteins, with glycoprotein B being one of the most abundant.
XX      CC Glycoprotein B from both HSV type I and type II are cross reactive. New
XX      CC vaccines comprising recombinantly produced glycosylated glycoprotein B
XX      CC that has a deletion of at least part of the transmembrane anchor region,
XX      CC in combination with a carrier and an adjuvant have been produced. The
XX      CC vaccines are used for immunising humans against HSV (HSV type 1 or
XX      CC HSV type 2) by vaccination before or after primary infection with HSV.
XX      SQ
XX      Sequence 3471 BP; 639 A; 1201 C; 1069 G; 562 T; 0 other;
XX
Query Match          100.0%; Score 20; DB 18; Length 3471;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TCACACCGTCAGCACCCTTC 20
DB      2290 TCACACCGTCAGCACCCTTC 2309

RESULT 12
ID      AAN80907
XX      AAN80907 standard; DNA; 3472 BP.
XX      AC
XX      AAN80907;
XX      DT
XX      25-MAR-2003 (updated)
XX      DT
XX      09-MAR-1992 (first entry)
XX      DE
XX      Sequence of Herpes simplex virus (HSV) glycoprotein B (GB) 1 DNA.
XX      KM
XX      Vaccine; herpes simplex virus; therapy; ss.
XX      OS
XX      Herpes simplex virus.
XX      PH
XX      Key          Location/Qualifiers

```

```

FT      CDS                               /note="putative"
FT      309..3023
FT      /*tag= a
FT      MO8802634-A.
XX      PD
XX      21-APR-1988.
XX      PF
XX      20-OCT-1987; 87WO-US02709.
XX      PR
XX      20-JUL-1987; 87US-0079605.
XX      PR
XX      20-OCT-1986; 86US-0921213.
XX      PA
XX      (CHIR ) CHIRON CORP.
XX      PI
XX      Burke LR, Pachl C, Valenzuela P;
XX      WPI; 1988-119368/17.
XX      DR
XX      P-PSDB; AAP80914.
XX      PT
XX      Vaccine for treatment of herpes simplex virus - contains
XX      recombinant HSV glyco:proteins B and D
XX      PS
XX      Example; Fig 4; 71pp; English.
XX      CC
XX      Prepn. of recombinant GB and GD is described in WO88504587. The
XX      CC amino acid sequence and DNA sequence for GB 1 presented in AAP80914
XX      CC and AAN80907 differ from that originally presented in Table 1 of
XX      CC International Publication No. WO 85/04587. The DNA sequence in Table
XX      CC 1 contains an error in that an additional nucleotide (G) is listed
XX      CC at position 607 which resulted in a shift in reading frame relative
XX      CC to AAN80907 from which this nucleotide has been deleted.
XX      CC (Updated on 25-MAR-2003 to correct PA field.)
XX      SQ
XX      Sequence 3472 BP; 639 A; 1202 C; 1068 G; 563 T; 0 other;
XX
Query Match          100.0%; Score 20; DB 9; Length 3472;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY      1 TCACACCGTCAGCACCCTTC 20
DB      2290 TCACACCGTCAGCACCCTTC 2309

RESULT 13
ID      AAQ48496
XX      AAQ48496 standard; DNA; 3472 BP.
XX      AC
XX      AAQ48496;
XX      DT
XX      25-MAR-2003 (updated)
XX      DT
XX      29-MAR-1994 (first entry)
XX      DE
XX      Glycoprotein B (GB1).
XX      KM
XX      Glycoprotein; GB1; GB2; herpes simplex virus; HSV-1; HSV-2;
XX      KM flanking; initiation; termination; transcription; translation;
XX      KM regulatory sequence; ss.
XX      OS
XX      Herpes simplex virus 1 strain Patcon.
XX      PH
XX      Key          Location/Qualifiers
XX      FT
XX      CDS          309..3023
XX      FT          /*tag= a
XX      PN
XX      US5244792-A.
XX      PD
XX      14-SEP-1993.
XX      PF
XX      20-SEP-1990; 90US-0587179.
XX      PR
XX      06-APR-1984; 84US-0597784.
XX      PR
XX      20-OCT-1986; 86US-0921730.

```

PR 20-SEP-1990; 90US-0587179.  
XX  
XX (CHIR ) CHIRON CORP.  
PI  
PI Burke RL, Pachl C, Valenzuela PDT;  
XX  
XX WPI: 1993-302641/38.  
DR P-PSDB; AAR41778.  
XX  
XX DNA construct for expressing HSV glycosylated polypeptide -  
PT useful for vaccinating against HSV-1 and -2 infections  
XX  
XX Disclosure; Fig 4; 33pp; English.  
XX  
XX DNA constructs for expressing a glycosylated polypeptide in a  
CC eukaryotic host cell comprises (1) an oligonucleotide sequence (OS)  
CC free of natural flanking sequences, encoding glycoprotein B (GB)  
CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating  
CC transcriptional and translational regulatory sequences flanking OS,  
CC at least one of these sequences not being from HSV. The HSV-1 and  
CC HSV-2 sequences are given in AAQ48496 and AAQ48497 respectively.  
CC (Updated on 25-MAR-2003 to correct PF field.)  
XX  
SQ Sequence 3472 BP; 639 A; 1202 C; 1069 G; 562 T; 0 other;  
Query Match 100.0%; Score 20; DB 14; Length 3472;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCAACGCTGACACCTTC 20  
DB 2290 TCACCAACGCTGACACCTTC 2309  
RESULT 14  
AAQ48497 standard; DNA; 3472 BP.  
XX  
XX AAQ48497;  
AC  
XX  
XX 25-MAR-2003 (updated)  
DT 29-MAR-1994 (first entry)  
XX  
XX Glycoprotein B (GB2).  
DE  
XX  
XX Glycoprotein; GB1; GB2; herpes simplex virus; HSV-1; HSV-2;  
KM flanking; initiation; termination; transcription; translation;  
KM regulatory sequence; ss.  
XX  
XX Herpes simplex virus 2 strain 333.  
OS  
XX  
XX Key Location/Qualifiers  
FH CDS 309..3023  
FT /\*tag= a  
FT  
XX  
XX US5244792-A.  
PN  
XX  
XX 14-SEP-1993.  
PD  
XX  
XX 20-SEP-1990; 90US-0587179.  
PF  
XX  
XX 06-APR-1984; 84US-0597784.  
PR 20-OCT-1986; 86US-0921730.  
PR 20-SEP-1990; 90US-0587179.  
PR 20-SEP-1990; 90US-0587179.  
XX  
XX (CHIR ) CHIRON CORP.  
PA  
XX  
XX Burke RL, Pachl C, Valenzuela PDT;  
PI  
XX  
XX WPI: 1993-302641/38.  
DR P-PSDB; AAR41779.  
XX  
XX DNA construct for expressing HSV glycosylated polypeptide -  
PT

PT useful for vaccinating against HSV-1 and -2 infections  
XX  
XX Disclosure; Fig 4; 33pp; English.  
XX  
XX DNA constructs for expressing a glycosylated polypeptide in a  
CC eukaryotic host cell comprises (1) an oligonucleotide sequence (OS)  
CC free of natural flanking sequences, encoding glycoprotein B (GB)  
CC of HSV, or its fragments, and (2) 5' initiating and 3' terminating  
CC transcriptional and translational regulatory sequences flanking OS,  
CC at least one of these sequences not being from HSV. The HSV-1 and  
CC HSV-2 sequences are given in AAQ48496 and AAQ48497 respectively.  
CC (Updated on 25-MAR-2003 to correct PF field.)  
XX  
SQ Sequence 3472 BP; 661 A; 1193 C; 1053 G; 565 T; 0 other;  
Query Match 100.0%; Score 20; DB 14; Length 3472;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCAACGCTGACACCTTC 20  
DB 2299 TCACCAACGCTGACACCTTC 2318  
RESULT 15  
AAT93651  
ID AAT93651 standard; DNA; 3473 BP.  
XX  
XX AAT93651;  
AC  
XX  
XX 26-FEB-1998 (first entry)  
DE  
XX  
XX Herpes simplex virus type 2 glycoprotein B encoding DNA.  
DE  
XX  
XX Herpes simplex virus; HSV; type I; type II; glycoprotein B; vaccine;  
KM membrane protein; virus-specific glycoprotein;  
KM transmembrane anchor region; ss.  
XX  
XX Herpes simplex virus type 2.  
OS  
XX  
XX Key Location/Qualifiers  
FH CAAAT\_signal 55..62  
FT /\*tag= a  
FT /note= "putative"  
FT TATA\_signal 125..131  
FT /\*tag= b  
FT /note= "putative"  
FT CDS 309..3023  
FT /\*tag= c  
FT polyA\_signal 2744..2751  
FT /\*tag= d  
FT /note= "putative"  
FT  
XX  
XX US5648079-A.  
PN  
XX  
XX 15-JUL-1997.  
PD  
XX  
XX 08-DEC-1994; 94US-0351875.  
PF  
XX  
XX 20-OCT-1986; 86US-0921730.  
PR 06-APR-1984; 84US-0597784.  
PR 20-SEP-1990; 90US-0587179.  
PR 21-DEC-1992; 92US-0993415.  
PR 18-OCT-1993; 93US-0138717.  
XX  
XX (CHIR ) CHIRON CORP.  
PA  
XX  
XX Burke RL, Pachl C, Valenzuela PDT;  
PI  
XX  
XX WPI: 1997-372022/34.  
DR P-PSDB; AAW34553.  
XX  
XX Vaccines against herpes simplex virus infection - containing  
PT

PT recombinant HSV glycoprotein B  
XX  
PS Disclosure; Fig 4A-F; 33pp; English.  
XX  
CC The present sequence is from the Herpes simplex virus (HSV) type 2  
CC strain 333 and encodes a glycoprotein B. HSV is a double stranded  
CC virus packaged within an icosahedral nucleocapsid enveloped within a  
CC membrane. The membrane includes a number of virus-specific  
CC glycoproteins, with glycoprotein B being one of the most abundant.  
CC glycoprotein B from both HSV type I and type II are cross reactive. New  
CC vaccines comprising recombinantly produced glycosylated glycoprotein B  
CC that has a deletion of at least part of the transmembrane anchor region,  
CC in combination with a carrier and an adjuvant have been produced. The  
CC vaccines are used for immunising humans against HSV (HSV type 1 or  
CC HSV type 2) by vaccination before or after primary infection with HSV.  
XX  
SQ Sequence 3473 BP; 661 A; 1195 C; 1053 G; 564 T; 0 other;  
Query Match 100.0%; Score 20; DB 18; Length 3473;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACGACCGTCAGCACCTTC 20  
|||  
DB 2299 TCACGACCGTCAGCACCTTC 2318

Search completed: September 16, 2003, 18:36:57  
Job time : 102.198 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 18:24:35 ; Search time 25.5446 Seconds  
(without alignments)  
345.579 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20  
Sequence: 1 tcaccaccgtcagcacccttc 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA:

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/CTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	2713	3	US-08-804-439A-13 Sequence 13, Appl
2	20	100.0	2713	3	US-08-720-229-13 Sequence 13, Appl
3	20	100.0	3472	6	5244792-1 Patent No. 5244792
4	20	100.0	3472	6	5244792-2 Patent No. 5244792
5	18.4	92.0	2088	1	US-08-164-839-7 Sequence 7, Appl
6	18.4	92.0	2088	1	US-08-583-799-7 Sequence 7, Appl
7	18.4	92.0	2921	1	US-08-164-839-9 Sequence 9, Appl
8	18.4	92.0	2921	1	US-08-583-799-9 Sequence 9, Appl
9	18.4	92.0	34094	4	US-09-292-034-1 Sequence 1, Appl
10	16.8	84.0	788	4	US-09-142-623-10 Sequence 10, Appl
11	16.8	84.0	1905	4	US-09-142-623-2 Sequence 2, Appl
12	16.8	84.0	4403765	3	US-09-103-840A-2 Sequence 2, Appl
13	16.8	84.0	4411529	3	US-09-103-840A-1 Sequence 1, Appl
14	16.4	82.0	2508	4	US-09-252-991A-793 Sequence 793, App
15	16.4	82.0	2577	4	US-09-252-991A-828 Sequence 828, App
16	15.8	79.0	1586	4	US-09-673-018-1 Sequence 1, Appl
17	15.8	79.0	2124	4	US-09-252-991A-12830 Sequence 12830, A
18	15.8	79.0	2454	3	US-08-890-615-1 Sequence 1, Appl
19	15.8	79.0	2454	3	US-09-246-290A-1 Sequence 1, Appl
20	15.8	79.0	2475	3	US-09-045-185-1 Sequence 1, Appl
21	15.8	79.0	2475	3	US-09-045-185-3 Sequence 3, Appl
22	15.4	77.0	264	4	US-09-252-991A-12642 Sequence 12642, A
23	15.4	77.0	1116	4	US-09-252-991A-13319 Sequence 13319, A
24	15.4	77.0	1491	2	US-08-941-647A-4 Sequence 4, Appl
25	15.4	77.0	1505	4	US-09-142-481-1 Sequence 1, Appl
26	15.4	77.0	2175	4	US-08-914-999-3 Sequence 3, Appl
27	15.4	77.0	2235	4	US-09-252-991A-12994 Sequence 12994, A

28	15.4	77.0	3486	4	US-09-252-991A-12477 Sequence 12477, A
29	15.4	77.0	25165	4	US-09-453-702B-39 Sequence 39, Appl
30	15.4	77.0	169998	4	US-09-676-610B-24 Sequence 24, Appl
31	15.4	77.0	197496	4	US-09-877-1177A-10 Sequence 10, Appl
32	15.4	77.0	4403765	3	US-09-103-840A-2 Sequence 2, Appl
33	15.4	77.0	4411529	3	US-09-103-840A-1 Sequence 1, Appl
34	15.2	76.0	36	4	US-09-689-291A-9 Sequence 9, Appl
35	15.2	76.0	100	4	US-09-298-886-20 Sequence 20, Appl
36	15.2	76.0	135	4	US-08-148-708-4 Sequence 4, Appl
37	15.2	76.0	291	4	US-09-313-294A-7433 Sequence 7433, Ap
38	15.2	76.0	294	4	US-09-313-294A-6329 Sequence 6329, Ap
39	15.2	76.0	296	4	US-09-313-294A-6788 Sequence 6788, Ap
40	15.2	76.0	377	2	US-08-332-766A-1 Sequence 1, Appl
41	15.2	76.0	420	4	US-09-262-537-22 Sequence 2369, Ap
42	15.2	76.0	429	4	US-09-262-537-22 Sequence 22, Appl
43	15.2	76.0	435	4	US-09-252-991A-2833 Sequence 2833, Ap
44	15.2	76.0	461	4	US-09-615-192A-375 Sequence 375, App
45	15.2	76.0	497	4	US-09-439-554-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-804-439A-13  
; Sequence 13, Application US/08804439A  
; Patent No. 6015565  
; GENERAL INFORMATION:  
; APPLICANT: Rose, Timothy M.  
; APPLICANT: Bosch, Martin L.  
; APPLICANT: Strand, Kurt  
; TITLE OF INVENTION: GLYCOPROTEIN B OF THE RHV/KSHV  
; TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES  
; NUMBER OF SEQUENCES: 113  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSES: Fish & Richardson P.C.  
; STREET: 4225 Executive Square, Ste 1400  
; CITY: La Jolla  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/804,439A  
; FILING DATE: February 21, 1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haile, Lisa A. 38,347  
; REGISTRATION NUMBER: 09176/004001  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 678-5070  
; TELEFAX: (619) 678-5099  
; INDEX:  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2713 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-804-439A-13  
Query Match 100.0%; Score 20; DB 3; Length 2713;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 1 TCACCACCGTCGACACCTTC 20  
Db 1988 TCACCACCGTCGACACCTTC 2007

RESULT 2  
US-08-720-229-13

Sequence 13, Application US/08720229  
Patent No. 6022542

GENERAL INFORMATION:

APPLICANT: Rose, Timothy M.

APPLICANT: Boesch, Marnix L.

APPLICANT: Strand, Kurt

TITLE OF INVENTION: GLYCOPROTEIN B OF THE HHV/KSHV

TITLE OF INVENTION: SUBFAMILY OF HERPES VIRUSES

NUMBER OF SEQUENCES: 100

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/720,229

FILING DATE: 26-SEP-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Schiffe, J. Michael

REGISTRATION NUMBER: 40,253

REFERENCE/DOCKET NUMBER: 29938-20002.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 2713 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-720-229-13

Query Match

Best Local Similarity 100.0%; Score 20; DB 3; Length 2713;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20

DB 1988 TCACCAACCTGACGACCTTC 2007

RESULT 3

5244792-1

Patent No. 5244792

APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.

TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYCOPROTEIN

NUMBER OF SEQUENCES: 19

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/587,179

FILING DATE: 20-SEP-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 921,730

FILING DATE: 20-OCT-1986

APPLICATION NUMBER: 597,784

FILING DATE: 06-APR-1984

SEQ ID NO:1:

LENGTH: 3472

5244792-1

Query Match

Best Local Similarity 100.0%; Score 20; DB 6; Length 3472;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20

DB 2290 TCACCAACCTGACGACCTTC 2309

RESULT 4

5244792-2

Patent No. 5244792

APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.

TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYCOPROTEIN

NUMBER OF SEQUENCES: 19

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/587,179

FILING DATE: 20-SEP-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 921,730

FILING DATE: 20-OCT-1986

APPLICATION NUMBER: 597,784

FILING DATE: 06-APR-1984

SEQ ID NO:2:

LENGTH: 3472

5244792-2

Query Match

Best Local Similarity 100.0%; Score 20; DB 6; Length 3472;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20

DB 2299 TCACCAACCTGACGACCTTC 2318

RESULT 5

US-08-164-839-7

Sequence 7, Application US/08164839

Patent No. 5514573

GENERAL INFORMATION:

APPLICANT: YASUEDA, HISASHI

APPLICANT: NAKANISHI, KAZUO

APPLICANT: MOTOKI, MASAO

APPLICANT: NAGASE, KAZUO

APPLICANT: MATSUI, HIROSHI

TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/164,839

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/004,729

FILING DATE: 14-JAN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5514573man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-599-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)412-3000  
TELEFAX: (703)413-2220  
TELEFAX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2088 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Theragra chalcogramma  
TISSUE TYPE: liver  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2085  
US-08-164-839-7

Query Match 92.0%; Score 18.4; DB 1; Length 2088;  
Best Local Similarity 95.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
Db 1802 TCACCACCGTCATCACCTTC 1821

RESULT 6  
US-08-583-799-7  
Sequence 7, Application US/08583799  
Patent No. 5607849  
GENERAL INFORMATION:  
APPLICANT: YASUEDA, HISASHI  
APPLICANT: NAKANISHI, KAZUO  
APPLICANT: MOTOKI, MASAO  
APPLICANT: NAGASE, KAZUO  
APPLICANT: MATSUI, HIROSHI  
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED  
TITLE OF INVENTION: FROM FISH  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/583,799  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/004,729  
FILING DATE: 14-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5607849man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-599-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)413-2220  
TELEFAX: (703)413-2220  
TELEFAX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2088 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Theragra chalcogramma  
TISSUE TYPE: liver  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2085  
US-08-583-799-7

Query Match 92.0%; Score 18.4; DB 1; Length 2088;  
Best Local Similarity 95.0%; Pred. No. 27;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
Db 1802 TCACCACCGTCATCACCTTC 1821

RESULT 7  
US-08-164-839-9  
Sequence 9, Application US/08164839  
Patent No. 5514573  
GENERAL INFORMATION:  
APPLICANT: YASUEDA, HISASHI  
APPLICANT: NAKANISHI, KAZUO  
APPLICANT: MOTOKI, MASAO  
APPLICANT: NAGASE, KAZUO  
APPLICANT: MATSUI, HIROSHI  
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED  
TITLE OF INVENTION: FROM FISH  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/164,839  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/004,729  
FILING DATE: 14-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5514573man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 10-599-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)413-2220  
TELEFAX: (703)413-2220  
TELEFAX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2921 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Theragra chalcogramma  
TISSUE TYPE: liver  
FEATURE:

NAME/KEY: CDS  
LOCATION: 32..2122  
US-08-164-839-9

Query Match 92.0%; Score 18.4; DB 1; Length 2921;  
Best Local Similarity 95.0%; Pred. No. 28;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
Db 1836 TCACCACCGTCATCACCCTTC 1855

## RESULT 8

US-08-583-799-9  
Sequence 9, Application US/08583799

PATENT No. 5607849  
GENERAL INFORMATION:  
APPLICANT: YASUEDA, HISASHI  
APPLICANT: NAKANISHI, KAZUO  
APPLICANT: MOTOKI, MASAO  
APPLICANT: NAGASE, KAZUO  
APPLICANT: MATSUI, HIROSHI  
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED  
TITLE OF INVENTION: FROM FISH  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 Jefferson Davis Highway, Fourth Floor  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/583,799

## FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/004,729

FILING DATE: 14-JAN-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5607849man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-599-0

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)412-3000

TELEFAX: (703)413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2921 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORGANISM: Theragra chalcogramma

TISSUE TYPE: liver

FEATURE:  
NAME/KEY: CDS  
LOCATION: 32..2122

US-08-583-799-9

Qy 1 TCACCACCGTCAGCACCTTC 20  
Db 1836 TCACCACCGTCATCACCCTTC 1855

## RESULT 9

US-09-292-034-1/c  
Sequence 1, Application US/09292034

PATENT No. 6492343  
GENERAL INFORMATION:  
APPLICANT: Reddy, P. Seethidhar  
APPLICANT: Bablu, Lorne  
APPLICANT: Bablu, Lorne  
TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME  
FILE REFERENCE: 29310202400  
CURRENT APPLICATION NUMBER: US/09/292,034  
CURRENT FILING DATE: 1999-04-14  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 34094  
TYPE: DNA  
ORGANISM: Porcine Adenovirus Type 3  
FEATURE:  
US-09-292-034-1

Query Match 92.0%; Score 18.4; DB 4; Length 34094;  
Best Local Similarity 95.0%; Pred. No. 31;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCACCACCGTCAGCACCTTC 20  
Db 11980 TCACCACCGTCAGCACCTTC 11961

## RESULT 10

US-09-142-623-10/c  
Sequence 10, Application US/09142623

PATENT No. 6337201

GENERAL INFORMATION:

APPLICANT: Koji YANAI et al.

TITLE OF INVENTION: -FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF

TITLE OF INVENTION: ISOLATING -FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCING

TITLE OF INVENTION: -FRUCTOFURANOSIDASE, AND -FRUCTOFURANOSIDASE VARIANT

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., Suite 800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/142,623

FILING DATE: September 10, 1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Lee Cheng

REGISTRATION NUMBER: 40,949

REFERENCE/DOCKET NUMBER: 98-0989\*/LC (WMC)/144

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 788 base pairs



TYPE: Nucleic acid  
STRANDEDNESS: Double stranded  
TOPOLOGY: Linear  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Aspergillus niger ACE-2-1  
ORGANISM: (ATCC 20611)  
FEATURES:  
NAME/KEY: P CDS (partial amino acid sequence)  
LOCATION: 1..788  
IDENTIFICATION METHOD: E  
US-09-142-623-10

Query Match 84.0%; Score 16.8; DB 4; Length 788;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCACCACGCTGACGACCTTC 20  
DB 772 TCTCCACCGCTGACGACCTTC 753

## RESULT 11

US-09-142-623-2/c  
Sequence 2, Application US/09142623  
Patent No. 6337201

GENERAL INFORMATION:  
APPLICANT: KOJI YANAI et al.  
TITLE OF INVENTION: -FRUCTOFRANOSIDASE AND ITS GENE, METHOD OF  
TITLE OF INVENTION: ISOLATING -FRUCTOFRANOSIDASE GENE, SYSTEM FOR PRODUCING  
TITLE OF INVENTION: -FRUCTOFRANOSIDASE, AND -FRUCTOFRANOSIDASE VARIANT  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
STREET: 2033 K Street, N.W., Suite 800  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/142,623  
FILING DATE: September 10, 1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee Cheng  
REGISTRATION NUMBER: 40,949  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-721-8200  
TELEFAX: 202-721-8250  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1905 base pairs  
TYPE: Nucleic acid  
STRANDEDNESS: Double stranded  
TOPOLOGY: Linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Microorganism: Aspergillus niger ACE-2-1  
ORGANISM: (ATCC 20611)  
FEATURES:

NAME/KEY: mat peptide  
LOCATION: 1..1905  
IDENTIFICATION METHOD: E  
US-09-142-623-2

Query Match 84.0%; Score 16.8; DB 4; Length 1905;

Best Local Similarity 90.0%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCACCACGCTGACGACCTTC 20  
DB 1243 TCTCCACCGCTGACGACCTTC 1224

## RESULT 12

US-09-103-840A-2/c  
Sequence 2, Application US/09103840A  
Patent No. 6294328

GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURES:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2

Query Match 84.0%; Score 16.8; DB 3; Length 4403765;  
Best Local Similarity 90.0%; Pred. No. 84;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCACCACGCTGACGACCTTC 20  
DB 4389770 TCACCACGCTGACGACCTTC 4389751

## RESULT 13

US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294328

GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 84.0%; Score 16.8; DB 3; Length 4411529;  
Best Local Similarity 90.0%; Pred. No. 84;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TCACCACGCTGACGACCTTC 20  
DB 4397533 TCACCACGCTGACGACCTTC 4397514



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2003, 19:14:25 ; Search time 72.2772 Seconds  
(without alignments)  
680.630 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20  
Sequence: 1 tcaccaccgtcagcacccttc 20

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 1660708 seqs, 122959015 residues

Total number of hits satisfying chosen parameters: 3321416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCT08\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	154746	12 US-09-827-688-8	Sequence 8, Appli
2	18.4	92.0	34094	12 US-09-963-038A-1	Sequence 1, Appli
3	18.4	92.0	34094	14 US-10-199-550-1	Sequence 1, Appli
4	18.4	92.0	34094	15 US-10-245-603A-1	Sequence 1, Appli
5	16.8	84.0	84	9 US-09-864-761-31756	Sequence 31756, A
6	16.8	84.0	466	9 US-09-864-761-2089	Sequence 2089, A
7	16.8	84.0	498	9 US-09-864-761-15233	Sequence 15233, A
8	16.8	84.0	498	11 US-09-918-995-26919	Sequence 26919, A
9	16.8	84.0	585	13 US-10-027-632-227474	Sequence 227474, A
10	16.8	84.0	788	10 US-09-990-385-10	Sequence 10, Appli
11	16.8	84.0	1336	12 US-10-017-161-569	Sequence 569, App
12	16.8	84.0	1336	12 US-10-017-161-1545	Sequence 1545, App
13	16.8	84.0	1339	12 US-10-017-161-813	Sequence 813, App
14	16.8	84.0	1629	14 US-10-128-714-2340	Sequence 2340, App
15	16.8	84.0	1691	10 US-09-938-842A-3140	Sequence 3140, App
16	16.8	84.0	1813	14 US-10-128-714-1340	Sequence 1340, Ap

17	16.8	84.0	1896	14 US-10-128-714-7340	Sequence 7340, Ap
18	16.8	84.0	1905	10 US-09-990-385-2	Sequence 2, Appli
19	16.8	84.0	2099	14 US-10-128-714-6340	Sequence 6340, Ap
20	16.8	84.0	3812	14 US-10-128-714-340	Sequence 340, App
21	16.8	84.0	4099	14 US-10-128-714-5340	Sequence 5340, App
22	16.4	82.0	628	13 US-10-027-632-289727	Sequence 289727, A
23	16.4	82.0	628	13 US-10-027-632-289728	Sequence 289728, A
24	16.4	82.0	628	13 US-10-027-632-289729	Sequence 289729, A
25	16.4	82.0	927	9 US-09-815-242-6115	Sequence 6115, App
26	16.4	82.0	2394	9 US-09-815-242-4016	Sequence 4016, App
27	16	80.0	243	10 US-09-969-708-166	Sequence 166, App
28	16	80.0	878	13 US-10-027-632-166052	Sequence 166052, A
29	16	80.0	878	13 US-10-027-632-166053	Sequence 166053, A
30	16	80.0	1918	10 US-09-822-863-1	Sequence 1, Appli
31	16	80.0	1918	12 US-10-281-319-1	Sequence 1, Appli
32	16	80.0	27359	10 US-09-822-863-3	Sequence 3, Appli
33	16	80.0	27359	12 US-10-281-319-3	Sequence 3, Appli
34	15.8	79.0	195	9 US-09-864-761-20885	Sequence 20885, A
35	15.8	79.0	212	11 US-09-764-891-267	Sequence 267, App
36	15.8	79.0	273	10 US-09-738-626-3003	Sequence 3003, App
37	15.8	79.0	278	9 US-09-864-761-28656	Sequence 28656, A
38	15.8	79.0	289	9 US-09-764-878-203	Sequence 203, App
39	15.8	79.0	289	14 US-10-079-854-203	Sequence 203, App
40	15.8	79.0	318	13 US-10-040-739-741	Sequence 741, App
41	15.8	79.0	323	10 US-09-878-574-2422	Sequence 2422, App
42	15.8	79.0	393	10 US-09-878-574-1401	Sequence 1401, App
43	15.8	79.0	409	9 US-09-864-761-4131	Sequence 4131, App
44	15.8	79.0	501	9 US-09-867-550-1533	Sequence 1533, App
45	15.8	79.0	549	9 US-09-864-761-12076	Sequence 12076, A

## ALIGNMENTS

RESULT 1  
US-09-827-688-8/c  
; Sequence 8, Application US/09827688  
; Publication No. US20030165476A1  
; GENERAL INFORMATION:  
; APPLICANT: ORSON, FRANK  
; APPLICANT: KINSEY, BREMA  
; APPLICANT: BHOAL, BALBIR  
; TITLE OF INVENTION: MAGROGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION D  
; FILE REFERENCE: P01949US1/10004014  
; CURRENT APPLICATION NUMBER: US/09/827,688  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/195,680  
; PRIOR FILING DATE: 2000-04-07  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 154746  
; TYPE: DNA  
; ORGANISM: HERPESVIRUS 2  
US-09-827-688-8

Query Match 100.0%; Score 20; DB 12; Length 154746;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 54136 TCACCACCGTCAGCACCTTC 54117

RESULT 2  
US-09-963-038A-1/c  
; Sequence 1, Application US/09963038A  
; Publication No. US20030143200A1  
; GENERAL INFORMATION:  
; APPLICANT: TIKOO, Sureeh K.  
; TITLE OF INVENTION: PORCINE ADENOVIRUS E1 REGION

FILE REFERENCE: 293102003200  
CURRENT APPLICATION NUMBER: US/09/963,038A  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PASTESEQ for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 34094  
TYPE: DNA  
ORGANISM: Porcine Adenovirus Type 3  
US-09-963-038A-1

Query Match 92.0%; Score 18.4; DB 12; Length 34094;  
Best Local Similarity 95.0%; Pred. No. 49;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCAGCACCCTTC 20  
DB 11980 TCACCACTGTCAGCACCCTTC 11961

## RESULT 3

US-10-199-550-1/c  
Sequence 1, Application US/10199550  
Publication No. US20030099615A1  
GENERAL INFORMATION:  
APPLICANT: TIKOO, Suresh K.  
TITLE OF INVENTION: PORCINE ADENOVIRUS E1 AND E4 REGIONS  
FILE REFERENCE: 293102003220  
CURRENT APPLICATION NUMBER: US/10/199,550  
CURRENT FILING DATE: 2002-12-20  
PRIOR APPLICATION NUMBER: US 09/963,038  
PRIOR FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PasteSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 34094  
TYPE: DNA  
ORGANISM: Porcine Adenovirus Type 3  
US-10-199-550-1

Query Match 92.0%; Score 18.4; DB 14; Length 34094;  
Best Local Similarity 95.0%; Pred. No. 49;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCAGCACCCTTC 20  
DB 11980 TCACCACTGTCAGCACCCTTC 11961

## RESULT 4

US-10-245-603A-1/c  
Sequence 1, Application US/10245603A  
Publication No. US2003010187A1  
GENERAL INFORMATION:  
APPLICANT: REDDY, Police Seshidhar  
APPLICANT: TIKOO, Suresh Kumar  
APPLICANT: BABIUK, Lorne A.  
TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME  
FILE REFERENCE: 293102002410  
CURRENT APPLICATION NUMBER: US/10/245,603A  
CURRENT FILING DATE: 2002-12-10  
PRIOR APPLICATION NUMBER: US 09/292,034  
PRIOR FILING DATE: 1999-04-14  
PRIOR APPLICATION NUMBER: US 60/081,882  
PRIOR FILING DATE: 1998-04-15  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PasteSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 34094  
TYPE: DNA  
ORGANISM: Porcine Adenovirus Type 3  
FEATURE:  
US-10-245-603A-1

Query Match 92.0%; Score 18.4; DB 15; Length 34094;  
Best Local Similarity 95.0%; Pred. No. 49;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCAGCACCCTTC 20  
DB 11980 TCACCACTGTCAGCACCCTTC 11961

## RESULT 5

US-09-864-761-31756  
Sequence 31756, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmics-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 31756  
LENGTH: 84  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC008179.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.91  
OTHER INFORMATION: EST HUMAN HIT: BB891293.1, EVALUATE 4.00e-03  
OTHER INFORMATION: NT HIT: L27595.1, EVALUATE 1.00e-05  
US-09-864-761-31756

Query Match 84.0%; Score 16.8; DB 9; Length 84;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Cy 1 TCACCGCTGACGACCTTC 20  
Db 35 TCACCGCTGACGACCTTC 54

## RESULT 6

US-09-864-761-2089/C  
Sequence 2089, Application US/09864761  
Patent No. US2002048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmca-X-1  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 2089  
LENGTH: 466  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC009303.1  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.86  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
US-09-864-761-2089

Query Match 84.0%; Score 16.8; DB 9; Length 466;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Cy 1 TCACCGCTGACGACCTTC 20  
Db 265 TCACCGCTGACGACCTTC 246

## RESULT 7

US-09-864-761-15233  
Sequence 15233, Application US/09864761  
Patent No. US2002048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmca-X-1  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 15233  
LENGTH: 498  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC008179.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.91  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.86  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
US-09-864-761-15233

Query Match 84.0%; Score 16.8; DB 9; Length 498;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACCACTGTGACACCTTC 20  
|||||  
DB 376 TCACCACTGTGACACCTTC 395

RESULT 8  
US-09-918-995-26919/c  
; Sequence 26919, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: HySeq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26919  
; LENGTH: 498  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(498)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-26919

Query Match 84.0%; Score 16.8; DB 11; Length 498;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TCACCACTGTGACACCTTC 20  
|||||  
DB 437 TCACCACTGTGACACCTTC 418

RESULT 9  
US-10-027-632-227474  
; Sequence 227474, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 227474  
; LENGTH: 585  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-227474

Query Match 84.0%; Score 16.8; DB 13; Length 585;

Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TCACCACTGTGACACCTTC 20  
|||||  
DB 30 TCACCACTGTGACACCTTC 49

RESULT 10  
US-09-990-385-10/c  
; Sequence 10, Application US/09990385  
; Publication No. US20020192771A1  
; GENERAL INFORMATION:  
; APPLICANT: KOJI YANAI et al.  
; TITLE OF INVENTION: BETA-FRUCTOFURANOSIDASE AND ITS GENE, METHOD OF  
; ISOLATING BETA-FRUCTOFURANOSIDASE GENE, SYSTEM FOR PRODUCI  
; BETA-FRUCTOFURANOSIDASE, AND BETA-FRUCTOFURANOSIDASE VARI  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.  
; STREET: 2033 K Street, N.W., Suite 800  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20006  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/990,385  
; FILING DATE: 10-Sep-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/142,623  
; FILING DATE: September 10, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lee Cheng  
; REGISTRATION NUMBER: 40,949  
; REFERENCE/DOCKET NUMBER: 2001-1611  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-721-8200  
; TELEFAX: 202-721-8250  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 788 base pairs  
; TYPE: Nucleic acid  
; STRANDEDNESS: Double stranded  
; TOPOLOGY: Linear  
; ORIGINAL SOURCE:  
; ORGANISM: Microorganism: Aspergillus niger ACE-2-1  
; (ATCC 20611)  
; FEATURE:  
; NAME/KEY: P CDS (partial amino acid sequence)  
; LOCATION: 1..788  
; IDENTIFICATION METHOD: E  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-990-385-10

Query Match 84.0%; Score 16.8; DB 10; Length 788;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACCACTGTGACACCTTC 20  
|||||  
DB 772 TCACCACTGTGACACCTTC 753

RESULT 11  
US-10-017-161-569  
; Sequence 569, Application US/10017161  
; Publication No. US20030143668A1  
; GENERAL INFORMATION:

APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 084335/0152  
CURRENT APPLICATION NUMBER: US/10/017,161  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: JP 2001/246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2430  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 569  
LENGTH: 1336  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: source  
LOCATION: (1)..(1336)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201)..(482)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (515)..(645)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (785)..(1136)  
US-10-017-161-569

Query Match 84.0%; Score 16.8; DB 12; Length 1336;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 217 TCACCAGGTGACGACCTTC 236

## RESULT 12

US-10-017-161-1545  
Sequence 1545; Application US/10017161  
Publication No. US20030143668A1  
GENERAL INFORMATION:  
APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 084335/0152  
CURRENT APPLICATION NUMBER: US/10/017,161  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: JP 2001/246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2430  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1545  
LENGTH: 1336  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: source  
LOCATION: (1)..(1336)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201)..(598)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (785)..(1136)  
US-10-017-161-1545

Query Match 84.0%; Score 16.8; DB 12; Length 1336;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TCACCACCGTCAGCACCTTC 20  
DB 217 TCACCAGGTGACGACCTTC 236

## RESULT 13

US-10-017-161-813  
Sequence 813; Application US/10017161  
Publication No. US20030143668A1  
GENERAL INFORMATION:  
APPLICANT: SUMA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
FILE REFERENCE: 084335/0152  
CURRENT APPLICATION NUMBER: US/10/017,161  
CURRENT FILING DATE: 2002-12-18  
PRIOR APPLICATION NUMBER: JP 2001/246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2430  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 813  
LENGTH: 1339  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: source  
LOCATION: (1)..(1339)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201)..(1139)  
US-10-017-161-813

Query Match 84.0%; Score 16.8; DB 12; Length 1339;  
Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
DB 217 TCACCAGGTGACGACCTTC 236

## RESULT 14

US-10-128-714-2340  
Sequence 2340; Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wengdi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Ercoshkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128,714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285,697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287,066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-08  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2340  
 ; LENGTH: 1629  
 ; TYPE: DNA  
 ; ORGANISM: Aspergillus fumigatus  
 US-10-128-714-2340

Query Match 84.0%; Score 16.8; DB 14; Length 1629;  
 Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACGACCGTCAGACCTTC 20  
 |||||  
 DB 461 TCACGACCGTCAGACCTTC 480

RESULT 15  
 US-09-938-842A-3140/c  
 ; Sequence 3140, Application US/09938842A  
 ; Patent No. US20020160378A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Harper, Jeff  
 ; APPLICANT: Krep, Joel  
 ; APPLICANT: Wang, Xun  
 ; APPLICANT: Zhu, Tong  
 ; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 ; FILE REFERENCE: SCRIPI300-3  
 ; CURRENT APPLICATION NUMBER: US/09/938,842A  
 ; PRIOR FILING DATE: 2001-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/227,866  
 ; PRIOR FILING DATE: 2000-08-24  
 ; PRIOR APPLICATION NUMBER: US 60/264,647  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/300,111  
 ; PRIOR FILING DATE: 2001-06-22  
 ; NUMBER OF SEQ ID NOS: 5379  
 ; SEQ ID NO 3140  
 ; LENGTH: 1691  
 ; TYPE: DNA  
 ; ORGANISM: Arabidopsis thaliana  
 US-09-938-842A-3140

Query Match 84.0%; Score 16.8; DB 10; Length 1691;  
 Best Local Similarity 90.0%; Pred. No. 2.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACGACCGTCAGACCTTC 20  
 |||||  
 DB 362 TCACGACCGTCAGACCTTC 343

Search completed: September 16, 2003, 22:54:09  
 Job time : 74.2772 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using SW model

Run on: September 16, 2003, 18:22:25 ; Search time 1012.28 Seconds  
(without alignments)  
480.194 Million cell updates/sec

Title: US-09-594-065-1

Perfect score: 20

Sequence: 1 tcaccaccgtcagcaccctc 20

Scoring table: IDENTITY NUC  
Gapop 10-0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_dln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	794	28	BH706301 BOMH09TF
2	18.4	92.0	460	28	BH558118 BOMH09TF
3	18.4	92.0	471	28	B2201556 BOMH09TF
4	18.4	92.0	537	9	AU291667 AU291667

C 5	18.4	92.0	563	12	BM619819
C 6	18.4	92.0	574	12	BM575164
C 7	18.4	92.0	568	13	BU374492
C 8	18.4	92.0	625	12	BM635604
C 9	18.4	92.0	635	12	BM638755
C 10	18.4	92.0	658	28	BH689640
C 11	18.4	92.0	666	12	BM619759
C 12	18.4	92.0	684	12	BM586254
C 13	18.4	92.0	687	12	BM648540
C 14	18.4	92.0	704	28	B2158304
C 15	18.4	92.0	712	14	CD261546
C 16	18.4	92.0	755	13	BU374843
C 17	18.4	92.0	762	9	AL929750
C 18	18.4	92.0	774	29	B2244382
C 19	18.4	92.0	841	29	B2240124
C 20	18.4	92.0	863	29	B2260714
C 21	18.4	92.0	1106	28	B2198952
C 22	18	90.0	394	14	CA706900
C 23	17.4	87.0	138	28	BH198208
C 24	17.4	87.0	308	10	BE495478
C 25	17.4	87.0	309	9	AV549908
C 26	17.4	87.0	358	13	BU837718
C 27	17.4	87.0	368	14	CA607934
C 28	17.4	87.0	387	10	BF146112
C 29	17.4	87.0	405	13	BU822753
C 30	17.4	87.0	413	13	AD433830
C 31	17.4	87.0	413	13	BU982582
C 32	17.4	87.0	420	9	AU230462
C 33	17.4	87.0	423	14	CA712183
C 34	17.4	87.0	426	14	CA711350
C 35	17.4	87.0	429	14	CD055963
C 36	17.4	87.0	432	13	BU982723
C 37	17.4	87.0	441	10	BG417681
C 38	17.4	87.0	442	14	CD057939
C 39	17.4	87.0	452	13	BU980793
C 40	17.4	87.0	466	13	BU650160
C 41	17.4	87.0	471	28	AQ912575
C 42	17.4	87.0	474	9	AV388895
C 43	17.4	87.0	482	13	BQ462866
C 44	17.4	87.0	502	10	BF145641
C 45	17.4	87.0	511	9	AU239184

## ALIGNMENTS

RESULT 1  
LOCUS BH706301 794 bp DNA linear GSS 20-FEB-2002  
DEFINITION BOMH09TF BO.2.3 KB Brassica oleracea genomic clone BOMH09,  
genomic survey sequence.  
ACCESSION BH706301  
VERSION BH706301.1 GI:18790010  
KEYWORDS GSS.  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosid II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 794)  
Tom,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.  
A whole genome shotgun sequencing of Brassica oleracea  
unpublished  
Other-GSSs: BOMH09TF  
COMMENT Contact: Chris Town .

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
Location/Qualifiers  
1. 794  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO100DH3"  
/db\_xref="taxon:3712"  
/clone\_1lb="BO. 2.3 KB"  
/note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt using BstXI linkers"

BASE COUNT 240 a 134 c 200 g 220 t

ORIGIN

Query Match 100.0%; Score 20; DB 28; Length 794;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20  
|||||  
122 TCACCAACCTGACGACCTTC 208

Db

RESULT 2  
BH558118/c 460 bp DNA linear GSS 14-DEC-2001  
LOCUS  
DEFINITION  
BOHMH39TR BOH Brassica oleracea genomic clone BOHMH39, genomic survey sequence.  
ACCESSION  
BH558118 GI:117809898  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Brassica oleracea  
Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
1 (bases 1 to 460)  
Town, C.D., Van Aken, S., Utebäck, T., Koo, H. and Fraser, C.M.  
TITLE  
JOURNAL  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished  
Other\_GSSs: BOHMH39TF  
COMMENT  
Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

FEATURES  
source  
Location/Qualifiers  
1. 460  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO100DH3"  
/db\_xref="taxon:3712"  
/clone\_1lb="BOHMH39"  
/note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt using BstXI linkers"

BASE COUNT 117 a 92 c 126 g 125 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 28; Length 460;  
Best Local Similarity 95.0%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20  
|||||  
122 TCACCAACCTGACGACCTTC 103

Db

RESULT 3  
BZ201556/c

LOCUS  
DEFINITION  
CH230-399M1.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone  
CH230-399M1, genomic survey sequence.  
ACCESSION  
BZ201556  
VERSION  
BZ201556.1 GI:23859608  
KEYWORDS  
GSS.  
SOURCE  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Rattus.

REFERENCE  
1 (bases 1 to 471)  
Zhao, S., Shetty, V., Shatsman, S., Teegaye, G., Geer, K., Shvartsbeyn, A., Gebreyes, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.  
Rat BAC End Sequences from Library CHORI-230 MboI segment  
Unpublished  
Other\_GSSs: CH230-399M1.TV  
CONTACT: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pjejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or erting\_information.html). BAC end page: http://www.tigr.org/ld/bac\_ends/rat/bac\_end\_intro.html  
Plate: 399 row: M column: 1  
Seq primer: T7  
Class: BAC ends.

FEATURES  
source  
Location/Qualifiers  
1. 471  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-399M1"  
/sex="female"  
/cell\_type="Brain"  
/clone\_1lb="CHORI-230 Segment 2"  
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SNHsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 123 a 128 c 116 g 104 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 28; Length 471;  
Best Local Similarity 95.0%; Pred. No. 1.3e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCAACCTGACGACCTTC 20  
|||||  
171 TCACCAACCTGACGACCTTC 152

Db

RESULT 4  
AU291667  
LOCUS  
DEFINITION  
AU291667 zinnia cultured mesophyll cell equalized cDNA zinnia  
elegans cDNA clone Z6332, mRNA sequence.  
ACCESSION  
AU291667  
VERSION  
AU291667.1 GI:24252175  
KEYWORDS  
EST.  
SOURCE  
Zinnia elegans  
Zinnia elegans  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; Campanulids; Asterales; Asteraceae; Asteroideae; Heliantheae; Zinnia.

REFERENCE 1 (bases 1 to 537)  
 AUTHORS Demura,T., Tashiro,G., Horiguchi,G., Kishimoto,N., Kubo,M., Matsuda,N., Minami,A., Nagata-Hiwatashi,M., Nakamura,K., Okamura,Y., Sassa,N., Suzuki,S., Yazaki,J., Kikuchi,S. and Fukuda,H.  
 TITLE Visualization by comprehensive microarray analysis of gene expression programs during transdifferentiation of mesophyll cells into xylem cells  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (24), 15794-15799 (2002)  
 COMMENT Contact: Taku Demura  
 Morphogenesis Research Group  
 RIKEN Plant Science Center  
 1-7-22 Suehirocho, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9605  
 Fax: 81-45-503-9573  
 Email: demura@postman.riken.go.jp  
 This clone was obtained at our laboratory.  
 Seq primer: M13 reverse

FEATURES  
 source  
 1..537  
 /organism="Zinnia elegans"  
 /mol\_type="mRNA"  
 /culturvar="Canary bird"  
 /db\_xref="taxon:34245"  
 /clone="Z6332"  
 /tissue\_type="mesophyll cell"  
 /clone\_lib="zinnia cultured mesophyll cell equalized cDNA"  
 /note="Vector: pGEM-T easy; cultured in tracheary element differentiation-inductive medium"

BASE COUNT 177 a 158 c 90 g 105 t 7 others

ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 537;  
 Best Local Similarity 95.0%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
 |||||  
 Db 312 TCACCACCGTCGCGCACCTTC 331

RESULT 5  
 BM619819/c 563 bp mRNA linear EST 25-FEB-2002  
 LOCUS 17000687440937 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone  
 DEFINITION 19600449618554 5', mRNA sequence.  
 ACCESSION BM619819  
 VERSION BM619819.1 GI:18918237  
 KEYWORDS EST.  
 ORGANISM Anopheles gambiae (African malaria mosquito)  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anophelinae.  
 1 (bases 1 to 563)  
 Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
 TITLE Celera Anopheles gambiae EST project  
 JOURNAL Unpublished  
 COMMENT Contact: Holt R.A.  
 Celera Genomics  
 45 W. Gude Dr., Rockville, MD 20850, USA  
 Tel: 2404533151  
 Fax: 2404534580  
 Email: HoltR@celera.com  
 Plate: NU01004N8 row: H column: 08  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1..563  
 /organism="Anopheles gambiae"  
 /mol\_type="mRNA"  
 /strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
 /db\_xref="taxon:7165"

/clone="19600449618554"  
 /dev\_stage="Adult"  
 /lab\_host="DH10b"  
 /clone\_lib="A.Gam.ad.cDNA1"  
 /note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 160 a 155 c 146 g 102 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 563;  
 Best Local Similarity 95.0%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20  
 |||||  
 Db 128 TCACCACCGTCAGCACCTTC 109

RESULT 6  
 BM575164/c 568 bp mRNA linear EST 22-FEB-2002  
 LOCUS 17000659210616 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone  
 DEFINITION 19600447014826 5', mRNA sequence.  
 ACCESSION BM575164  
 VERSION BM575164.1 GI:18863631  
 KEYWORDS EST.  
 ORGANISM Anopheles gambiae (African malaria mosquito)  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides; Anophelinae.  
 1 (bases 1 to 568)  
 Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
 TITLE Celera Anopheles gambiae EST project  
 JOURNAL Unpublished  
 COMMENT Contact: Holt R.A.  
 Celera Genomics  
 45 W. Gude Dr., Rockville, MD 20850, USA  
 Tel: 2404533151  
 Fax: 2404534580  
 Email: HoltR@celera.com  
 Plate: NU01003ST row: G column: 09  
 Seq primer: M13 Reverse.

FEATURES  
 source  
 1..568  
 /organism="Anopheles gambiae"  
 /mol\_type="mRNA"  
 /strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
 /db\_xref="taxon:7165"  
 /clone="19600447014826"  
 /dev\_stage="Adult"  
 /lab\_host="DH10b"  
 /clone\_lib="A.Gam.ad.cDNA.blood1"  
 /note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 175 a 139 c 136 g 118 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 568;  
 Best Local Similarity 95.0%; Pred. No. 1.3e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACCGTCAGCACCTTC 20

Db 61 TCACCTCCGTCAGCACCTTC 42

RESULT	LOCUS	DEFINITION	ACCSSION	VERSION	KEYWORDS	SOURCE	ORGANISM
3	BU374492	574 bp. mRNA	BU374492	EST	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus
		linear	BU374492	EST	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus
		EST 28-NOV-2002	BU374492.1	GI:25882493	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus
		sequence.	BU374492	EST	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus
		60381877/1 CSECHN74	BU374492	EST	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus
		Gallus gallus CDNA	BU374492	EST	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus
		clone CHEST80420 5', mRNA	BU374492	EST	Gallus gallus (chicken)	Gallus gallus	Chicken, Gallus gallus

```

BASE COUNT      152 a       140 c       170 g       112 t
ORIGIN
Query Match          92.0% ; Score 18.4 ; DB 13 ; Length 574 ;
Best Local Similarity 95.0% ; Pred. No. 1.4e+03 ;
Matches   19; Conservative    0; Mismatches    1; Indels     0; Gaps     0;

Oy      1 TCACCACCGTCAGCAGCCCTTC 20
        |||||
Db      211 TCACCAACGTGACGACACTTC 230

RESULT 8
BM635604/c      BM635604      625 bp      mRNA      linear      EST 26-FEB-2002
LOCUS           17000667558989 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
DEFINITION

```

19600449658640 5', mRNA sequence  
 ACCESSION BM635604  
 VERSION BM635604.1 GI:18935115

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Anopheles gambiae (African malaria mosquito)	Anopheles gambiae	1 (pages 1 to 625)	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Krafc, C.L., Charlabb			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			R., Collins, P.H., Venter, J.C. and Hoffman, S.T.			
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;			Celera Anopheles gambiae EST project			
Anophales.			Unpublished			
			Contact: Holt R.A.			

```

ORIGIN
BASE COUNT      140 a      167 c      180 g      138 t
Query Match      92.0% ; Score 18.4 ; DB 12 ; Length 625 ;
Best Local Similarity 95.0% ; Pred. No. 1.4e+03 ;
Matches 19 ; Conservative 0 ; Mismatches 1 ; Indels 0 ; Gaps 0 ;

QY      1 TCACCACGTCAGACCTTC 20
        |||||
Db      307 TCACCACTGTGACACCTTC 288

RESULT 9
LOCUS      BM638755/c
DEFINITION 17000687228313 A.Gam.ad.cDNA1 Anopheles gambiae linear EST 26-FEB-2002
VERSION    19600449617587 5', mRNA sequence.
ACCESSION  BM638755
KEYWORDS   BM638755.1 GI:18938278
SOURCE     EST.
ORGANISM   Anopheles gambiae (African malaria mosquito)
            Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
            Anopheles.
REFERENCE  1 (bases 1 to 635)
            Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
            Celera Anopheles gambiae EST project
            Unpublished
            Contact: Holt R.A.
            Celera Genomics
            45 W. Gude Dr., Rockville, MD 20850, USA
            Tel.: 2404533151

```

FEATURES  
source  
Fax: 2404534580  
Email: HoltR@celera.com  
Plate: NU01004N8P row: P column: 01  
Seq primer: M13 Reverse.

Location/Qualifiers  
1. 635  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449617587"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_lib="A. Gam. ad. cDNA1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 164 a 145 c 221 g 105 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 635;  
Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCAGCACCTTC 20  
|||||  
Db 477 TCACCTCGTCAGCACCTTC 458

RESULT 10  
LOCUS BH689640 658 bp DNA linear GSS 19-FEB-2002  
DEFINITION BOMGX07R BO\_2\_3 KB Brassica oleracea genomic clone BOMGX07,  
genomic survey sequence.  
ACCESSION BH689640  
VERSION BH689640.1 GI:18760077  
KEYWORDS GSS  
SOURCE Brassica oleracea  
ORGANISM Brassica oleracea  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosid1 II; Brassicales; Brassicaceae; Brassica.  
1 (bases 1 to 658)  
TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished  
Other GSSs: BOMGX07TF  
COMMENT Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: 1R  
Classes: sheared ends.

FEATURES  
source  
Location/Qualifiers  
1. 658  
/organism="Brassica oleracea"  
/mol\_type="genomic DNA"  
/strain="TO100DH3"  
/db\_xref="taxon:3712"  
/clone="BOMGX07"  
/clone\_lib="BO\_2\_3 KB"  
/note="Vector: pHO51; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHO51 using BstXI linkers"

BASE COUNT 154 a 211 c 142 g 151 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 28; Length 658;  
Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCAGCACCTTC 20  
|||||  
Db 309 TCACCACTGTCAGCACCTTC 328

RESULT 11  
LOCUS BM619759/c 666 bp mRNA linear EST 25-FEB-2002  
DEFINITION 17000687440855 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone  
19600449618598 5', mRNA sequence.  
BM619759  
BM619759.1 GI:18918177

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Anopheles gambiae (African malaria mosquito)  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anopheles.  
1 (bases 1 to 666)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltR@celera.com  
Plate: NU01004N8P row: J column: 04  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 666  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449618598"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_lib="A.Gam.ad.cDNA1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

FEATURES  
source  
Location/Qualifiers  
1. 666  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449618598"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_lib="A.Gam.ad.cDNA1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

BASE COUNT 188 a 156 c 192 g 130 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 12; Length 666;  
Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCACTGTCAGCACCTTC 20  
|||||  
Db 229 TCACCTCGTCAGCACCTTC 210

RESULT 12  
LOCUS BM586254 684 bp mRNA linear EST 25-FEB-2002  
DEFINITION 17000687312069 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone  
19600449707650 5', mRNA sequence.  
BM586254  
BM586254.1 GI:18882115  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
Anopheles gambiae (African malaria mosquito)

ORGANISM  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

REFERENCE  
1 (bases 1 to 684)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU010049VI row: H column: 04  
Seq primer: M13 Reverse.  
Location/Qualifiers

FEATURES  
source  
1..684  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449707650"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_1lb="A.Gam.ad.cDNA.bloodi"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
hours after human blood feeding. cDNA inserts >500 bp  
cloned directionally into pSport 1. Not 1 site is 3'.  
Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)"

BASE COUNT  
182 a 158 c 232 g 112 t

ORIGIN  
Query Match 92.0%; Score 18.4; DB 12; Length 684;  
Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  
1 TCACCAACGTCAGCACCTTC 20  
|||||  
523 TCACCTCCGTCAGCACCTTC 504

RESULT 13  
BM648540 687 bp mRNA linear EST 26-FEB-2002  
LOCUS 17000687326166 A.Gam.ad.cDNA1 Anopheles gambiae CDNA clone  
DEFINITION 19600449658826 5', mRNA sequence.  
ACCESSION BM648540  
VERSION BM648540.1 GI:18948051  
KEYWORDS EST.  
SOURCE Anopheles gambiae (African malaria mosquito)  
ANALYSIS Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
1 (bases 1 to 687)  
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
Celera Anopheles gambiae EST project  
Unpublished  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: HoltRA@celera.com  
Plate: NU010049VI row: F column: 08  
Seq primer: M13 Reverse.  
Location/Qualifiers

FEATURES

source  
1..687  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449658826"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/clone\_1lb="A.Gam.ad.cDNA1"  
/note="Vector: pSport1; Site 1: SalI; Site 2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 1.  
cDNA inserts >500 bp cloned directionally into pSport 1.  
Not 1 site is 3'. Clones available through the Malaria  
Research and Reference Reagent Resource Center  
(www.malaria.mr4.org)."

BASE COUNT  
175 a 157 c 238 g 117 t

ORIGIN  
Query Match 92.0%; Score 18.4; DB 12; Length 687;  
Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  
1 TCACCAACGTCAGCACCTTC 20  
|||||  
546 TCACCTCCGTCAGCACCTTC 527

RESULT 14  
B2158304 704 bp DNA linear GSS 11-OCT-2002  
LOCUS CH230-38714.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone  
DEFINITION CH230-38714, genomic survey sequence.  
ACCESSION B2158304  
VERSION B2158304.1 GI:23799259  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ANALYSIS Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 704)  
Zhao,S., Shetty,U., Shatsman,S., Tsegaye,G., Geer,K., Shvartbeyn  
A., Gebregorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de  
Jong,P. and Fraser,C.M.  
Rat BAC End Sequences from Library CHORI-230 MboI segment  
Unpublished  
Other\_GSSs: CH230-38714.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact pieter de Jong (pjejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or eting\_information.htm). BAC end  
page: http://www.tigr.org/cdb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 387 row: I column: 4  
Seq primer: SP6  
Class: BAC ends.  
Location/Qualifiers

FEATURES  
source  
1..704  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-38714"  
/sex="Female"  
/cell\_type="Brain"

/clone\_lib="CHORI-230 Segment 2"  
 /note="Vector: pTARBACT.3; Site\_1: MboI; Site\_2: MboI;  
 CHORI-230 Rat (BN/SvNsd/MCW) BAC library produced by  
 Pieter de Jong"  
 BASE COUNT 156 a 171 c 204 g 173 t  
 ORIGIN

Query Match 92.0%; Score 18.4; DB 28; Length 704;  
 Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGCAGCTTC 20  
 |||||  
 Db 74 TCACCAACGTCAGCAGCTTC 93

## RESULT 15

CD261546 712 bp mRNA linear EST 23-MAY-2003  
 LOCUS pSMA012X015F.189662 SMA: Phytophthora sojae grown in synthetic  
 DEFINITION medium Phytophthora sojae CDNA clone SMA012015 5, mRNA sequence.

ACCESSION CD261546

VERSION CD261546.1 GI:31049373

KEYWORDS EST.

SOURCE Phytophthora sojae

ORGANISM Phytophthora sojae

Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;

Phytophthora.

1 (bases 1 to 712)

REFERENCE Tyler,B.M., Udelsom,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.  
 USDA-IRAFs: Expression of Phytophthora sojae genes during infection  
 and propagation

Unpublished

CONTACT: Tyler B

Journal

comment

Tyler lab

VBI  
 1880 Pratt Dr., Blacksburg, VA 24061, USA

Tel: 540-231-7318

Email: bmtyley@vt.edu

PCR Primers

FORWARD: BK reverse

Plate: 012 row: 0 column: 15

Seq primer: BK reverse

High quality sequence stop: 712.

Location/Qualifiers

1..712

/organism="Phytophthora sojae"

/mol\_type="mRNA"

/db\_xref="taxon:67593"

/clone="SMA012015"

/tissue\_type="mycelium"

/cell\_line="P6497"

/dev\_stage="mycelium"

/clone\_lib="SMA: Phytophthora sojae grown in synthetic

medium"

/note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT 172 a 203 c 213 g 124 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 712;  
 Best Local Similarity 95.0%; Pred. No. 1.4e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCACCAACGTCAGCAGCTTC 20  
 |||||  
 Db 693 TCACCAACGTCAGCAGCTTC 712

Search completed: September 16, 2003, 20:39:49  
 Job time : 1016.28 secs

BEST AVAILABLE COPY